

THE UNIVERSITY
OF ILLINOIS
LIBRARY

370
Ie6
no.14-17

EDUC. 1174
REF. 1174
Education

The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.

To renew call Telephone Center, 333-8400

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

OCT 18 1982

SEP 27 1982

MAILED

L161—O-1096

343. WOODROW, HERBERT. *Brightness and Dullness in Children.* Philadelphia: J. B. Lippincott and Company, 1919. 322 p.

There are various discussions of retardation and advancement, elimination, special classes, bright and dull pupils and experiments with bright pupils.

344. WOODS, ELIZABETH L. "Provision for the gifted child," *Educational Administration and Supervision*, 3:139-49, March, 1917.

A summary of provisions in a large number of cities showing that flexible promotion is very common, special classes or groups fairly so, and that practically all superintendents favor making some such provision.

345. YOUNG, ELLA FLAGG. "Grading and classification of pupils," *Addresses and Proceedings of the National Educational Association*, 32:83-86, 1893.

A rather general discussion.

346. "The Cambridge plan," *Journal of Education*, 75:375-76, April 4, 1912.

A brief description of this plan.

BULLETIN No. 17

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE PRESENT
STATUS OF WRITTEN EXAMINATIONS
AND SUGGESTIONS FOR THEIR
IMPROVEMENT

By

WALTER S. MONROE
Director, Bureau of Educational Research

Assisted by

LLOYD B. SOUDERS
Formerly Assistant in Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1923



TABLE OF CONTENTS

	PAGE
PREFACE.	5
CHAPTER I.—INTRODUCTION.	7
CHAPTER II.—SUMMARY OF CRITICISMS OF WRITTEN EXAMINATIONS.	9
Examinations yield inaccurate measures of school achievement.	9
Written examinations tend to encourage undesirable mental processes.	13
Passing the final examination an undesirable objective. . . .	14
Examinations injurious to health of students.	14
Time devoted by teachers to written examinations not profitably spent.	15
CHAPTER III.—PREPARATION AND ADMINISTRATION OF EXAMINATIONS IN HIGH SCHOOL.	16
The data collected.	16
Requirement of final examinations in Illinois high schools	16
Time devoted to written examinations.	19
Characteristics noted in marking examination papers. . . .	20
Weighting of questions.	22
Recognition of rate of work.	22
Methods of marking examination papers.	23
Directions to students concerning methods of work.	24
Recognition of a standard distribution in assigning grades to examination papers.	24
Relation of examination grades to final grades.	25
Summary	25

CHAPTER IV.—THE CONSTANT AND VARIABLE ERRORS IN EXAMINATION GRADES.....	27
Constant and variable errors of measurement.....	27
Magnitude of variable errors of measurement in standardized test scores and in examination grades.....	28
Methods employed in present investigation concerning reliability of written examinations.....	29
Data collected for investigation.....	30
Reliability of written examination grades and of standardized test scores.....	31
Conditions tending to produce variable errors of measurement in examination grades.....	37
Magnitude of constant errors of measurements in standardized test scores and in examination grades.....	40
Results of present investigation and of previous studies compared.....	41
Conclusion—relative accuracy of examination grades and of test scores.....	41
CHAPTER V.—THE CONTENT OF WRITTEN EXAMINATIONS.....	43
The data collected.....	43
Classification of questions.....	43
Relation of examination questions to educational objectives.....	46
CHAPTER VI.—THE IMPROVEMENT OF WRITTEN EXAMINATIONS.	48
Reduction of constant errors.....	50
Reduction of variable errors.....	54
Agreement of content of examinations with educational objectives.....	55
Simplification of administration of written examinations..	56
CHAPTER VII.—RULES FOR THE PREPARATION AND ADMINISTRATION OF WRITTEN EXAMINATIONS.....	63
APPENDIX.....	66


PREFACE

This bulletin reports the results of three extensive investigations relating to written examinations. These investigations were made by Mr. Souders under the direction of the Director of the Bureau of Educational Research. The tabulations and statistical calculations were made by Mr. Souders or by clerks working under his immediate direction. The preparation of the published report, however, is the work of the Director of the Bureau.

The Bureau of Educational Research wishes to acknowledge its indebtedness to the superintendents, principals, and teachers who cooperated by furnishing the necessary data. The data required in the study of reliability of written examinations necessitated considerable additional labor. Without their cooperation these investigations would not have been possible.

WALTER S. MONROE, *Director*.

November 1, 1923.



Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

PRESENT STATUS OF WRITTEN EXAMINATIONS AND SUGGESTIONS FOR THEIR IMPROVEMENT

CHAPTER I INTRODUCTION

Preparation and administration of written examinations important phases of the teacher's work. Written examinations, except in the few schools where they have been abolished, form a very important phase of the teacher's work, both because of the time devoted to their preparation and administration and of the significance attached to the measures which they yield. The final grades upon which promotion and the awarding of school honors depend are determined largely by final examinations and by written tests given during the school term. Altho standardized educational tests have become widely used during recent years, written examinations are still the most frequently used type of measuring instrument. This will probably always be true, particularly in the high school. Hence, we may expect that written examinations will occupy in the future as in the past, an important place in the work of our schools.

Need for more information concerning written examinations. There have been numerous investigations which showed that the marking of written examination papers is highly subjective—that is, different teachers tend to assign different marks to the same paper. With the exception of these studies relatively little precise information is available in regard to written examinations but a number of criticisms based upon experience and theoretical considerations have been made. As a result many teachers and other school officials have come to consider written examinations very inferior instruments and have abolished them in a number of schools.

A search through our educational literature, particularly textbooks, reveals an astonishing lack of information in regard to written examinations. Relatively little specific attention has been

given to their preparation and administration in our courses for the training of teachers. Inexperienced teachers have been left largely to their own devices in this important phase of their work. Careful inquiry and observation have indicated that there is a variety of practises with reference to the types of questions asked and to the administration of written examinations. Hence it appears that there is need for a comprehensive investigation of the present status of written examinations in order that a more intelligent estimate may be formed of their value in the process of education and that specific directions may be formulated in regard to their preparation and administration.

Purpose of this bulletin. It is the purpose of this bulletin to present (1) a brief summary of certain previous investigations relating to written examinations and also of the arguments for and against written examinations; (2) the results of three extensive investigations conducted by the Bureau of Educational Research, (a) the preparation and administration of written examinations in Illinois high schools, (b) the constant and variable errors in examination grades, and (c) the content of written examinations; and (3) suggestions for the improvement of written examinations. In the concluding chapter the author presents a list of rules in regard to the preparation and administration of written examinations.

CHAPTER II

SUMMARY OF CRITICISMS OF WRITTEN EXAMINATIONS¹

Plan of chapter. In this chapter the important criticisms of written examinations are briefly summarized. Following each criticism either a brief answer is given or a reference is made to a detailed discussion in a later chapter. By presenting both sides of the question in this way, it is hoped that the reader will be assisted in forming an intelligent estimate of the merits of written examinations.

I. Examinations yield inaccurate measures of school achievement. In support of this argument six points have been made.

1. The most important criticism relating to the accuracy of written examinations is that the marking of the papers is highly subjective. A large number of scientific investigations have yielded objective evidence that different teachers when working independently tend to assign widely varying marks to the same paper. One of the first studies of this type was by Starch and Elliott who found that the marks assigned to the same examination paper in Plane Geometry by 116 teachers ranged from 28 to 92 on the scale of 100 percent. The facts of such investigations as this can not be disputed but as we have no means of securing perfectly accurate measures of achievement, the question at issue concerns the relative rather than the absolute accuracy of the measurements secured. Facts may be misinterpreted. In Chapter IV we shall present evidence to show that when judged in relation to other means for measuring school achievement, written exami-

¹Starch, Daniel, and Elliott, E. C. "Reliability of grading high-school work in mathematics," *School Review*, 21:254-59, April, 1913.

Morton, Robert L. "The examination method of licensing teachers," *Educational Administration and Supervision*, 6:421, November, 1920.

Wood, Ben D. "Measurement of college work," *Educational Administration and Supervision*, 7:301-34, September, 1921.

Kelly, F. J. "Teachers' marking," *Teachers College Contributions to Education*, No. 66. New York: Teachers College, Columbia University, 1914.

nations yield relatively more accurate measures than generally supposed. In view of the additional information secured the subjectivity of written examinations loses much of its potency as a reason for their abolition.

2. The questions of ordinary examinations are usually not equal in difficulty and weighting by teachers is highly subjective.² It has been inferred that this condition tends to increase materially the inaccuracy of examination marks. Comparisons of weighted and non-weighted scores yielded by standardized tests have revealed that the errors introduced by disregarding the unequal difficulty of exercises or questions are not significant in most cases.³

3. It has been pointed out that frequently the content of written examinations is not in agreement with recognized educational objectives. Catch questions relating to trivial facts or worded in a misleading way have been cited as illustrations. Certain examination questions also have referred to items which had not been included in the course or at least had received only minor emphasis. Some evidence with reference to the justification of this criticism will be presented in Chapter V.

4. In most examinations the rate of work is neglected. The usual practise is to allow sufficient time for all pupils to finish or to base the mark only on the questions answered in the unfinished papers. Hence a student's examination grade is not influenced by the rate at which he answers the questions. It is easily possible to take into account the student's rate of work in determining the mark assigned to his examination paper. One plan is to set an examination of sufficient length so that all members of the class will be employed during the entire period. Another procedure is to have the student record the time when he finishes. In this way some weight can be given to his rate of work. This criticism is, however, a minor one. In some subjects the rate of work is an important consideration but in others, particularly those in which reasoning predominates in answering the questions, the neglect of the rate of work will affect the accuracy of the examination marks only slightly, if at all.

²Comin, Robert. "Teachers' estimates of the abilities of pupils," *School and Society*, 3:67-70, January 8, 1916.

³Charters, W. W. "Constructing a language and grammar scale," *Journal of Educational Research*, 1:249-58, April, 1920.

Monroe, Walter S. "The description of the performances of pupils on exercises of varying difficulty," *School and Society*, 15:341-43, March, 1922.

5. Written examinations are usually so short that they do not offer an adequate opportunity for a student to demonstrate his ability. This criticism is frequently expressed in the statement that it is unjust to base a student's standing for a semester or a year on an examination paper written during a brief examination period. When stated in this way the criticism refers to two issues between which there is failure to distinguish. The first is in regard to the weight allowed the examination grades in determining a student's final standing. This question of the weight given the final examination is discussed in a later chapter, but it may be said here that the usual practise in high schools is to count the written examination as one-third of the student's total grade. The second refers to the inaccuracy of the grade due to the limited opportunity which is given the student to demonstrate his ability. For practical reasons it is necessary that measurement of the total achievement for the term be based upon a sample. In general, increasing the scope of the examination will tend to increase the accuracy of the measures yielded. Some evidence with reference to the reliability of examination grades based upon short samples will be presented in Chapter IV. It is possible for a teacher to make examinations more comprehensive. This can be accomplished in part by exercising more care in the preparation of the questions. The "new examination" in which pupils are required to do little or no writing affords one means for covering a wide range of subject-matter in a brief period. This method of improving examinations will be discussed in Chapter VI.

The final point to be made with reference to the inaccuracy of examination marks refers to the distinction between a "score" which describes a pupil's performance on the examination and a "grade" which interprets this score with reference to a norm. Failure to recognize this distinction is primarily responsible for too high grading by some teachers and too low by others. Even the same teacher is likely to assign "high grades" on some examinations and "low grades" on others.

In order to understand how norms (standards) are used in connection with the grading of examination papers it is necessary to distinguish between "scores," or measures, and "grades," or marks. A "score" simply describes the performance which has been recorded on the examination paper. For example, a pupil may answer 55 percent of the questions correctly. In this case 55

is his "score." If a certain number of points or credits had been given for each question his score might be 129, or 91, or 217. A "grade" interprets this description with reference to certain norms. A "grade" indirectly describes a pupil's performance on an examination, but it tells also whether the performance is to be considered as above or below passing; whether the pupil is to receive the highest mark or the lowest mark or an average mark.

It is customary to describe the quality of examination papers in terms of the percent of questions answered correctly. For example, if an examination includes ten questions and a pupil answers seven of them correctly and an eighth one partially right, he is given a score of 75 percent, which is interpreted to mean that in the judgment of the examiner he has answered the questions 75 percent correctly. School marks or "grades" are also frequently expressed in terms of percents. Sometimes they are expressed in terms of letters or other symbols, but these in turn are defined in terms of percents. For example, the grade of "A" may be defined as being between 95 percent and 100 percent. Since both "scores" and "grades" are generally expressed in terms of percents, it is only natural that the two have been confused and that "scores" have been used as "grades."

A good illustration of their difference came to the writer recently. An examination in mathematics was given to nearly 1000 freshmen in one of our large universities. This examination may properly be described as "hard," considering the training which the students had received. One student made a score of 100. The lowest score was 12. The average was approximately 55. From the standpoint of the distribution of scores this was a "good examination." If it had been easier, so that any considerable number of pupils received scores of 100 percent, it would have been unsatisfactory. If it had been so "hard" that a considerable number of students made zero scores it would also have been defective. In both cases it would have failed to differentiate between some students who were not equal in ability. However, obviously an injustice would be done if a passing mark of 70 or 75 were adopted and all pupils having scores below this mark were given a grade of failure. The passing mark for this particular examination should be in the neighborhood of 40. If the "scores" are to be represented in terms of "grades" a "score" of 40 should be translated into a "grade" of 70 or whatever passing mark has been adopted by the institution.

The recognition of this distinction between "scores" and "grades" enables us to indicate the way in which subjective norms are implied in "grades." A "grade" is not a pure measure or description of the pupil's performance. It is rather an interpretation of the measure of his performance with reference to certain norms. When no distinction is made and "scores" are used as "grades," pupils will receive high "grades" if the examination is "easy;" and low ones if it is "hard." Thus, the difficulty of the examination is one factor in establishing the norms with reference to which the "scores" are interpreted when they are used as "grades." Severe marking will tend to set high norms. Only when the examination is of average or "standard" difficulty and the marking is average in severity do "scores" and "grades" become identical in magnitude. Since the norms are established by the difficulty of the examination and the severity of the scoring, they must be subjective. In the investigations of the marking of examination papers it was shown that teachers varied widely in their judgments concerning the worth of examination papers. There is no reason to expect that they would agree more closely in estimating the difficulty of examinations. Hence, norms which depend upon teachers' estimates of the questions appropriate for examinations and upon their marking of the papers must be considered subjective. It is possible to increase greatly the objectivity of these norms and the first requirement is to recognize the distinction between "scores" and "grades." (See page 38 for a further consideration of this topic.)

Summary of inaccuracy of examination marks. From the preceding discussion examination marks are, without doubt, shown to be far from accurate measures of school achievement. However, it does not necessarily follow that the errors involved are of sufficient magnitude to justify the abolishment of written examinations. In the writer's belief the greatest benefit will come from making an intelligent inquiry into the nature of these errors and from taking steps to reduce them to the lowest magnitude.

II. Written examinations tend to encourage undesirable mental processes. Many critics have claimed that most examinations, particularly those given at the end of a course, tend to encourage "cramming." The assertion is made that many students do little or no studying until near the close of the term. Then by the process of "cramming" they are able to pass the final ex-

amination and attain a relatively high standing in the course. This criticism assumes that "cramming" is an undesirable mental process and that final examinations are responsible for its occurrence. The undesirable feature is the neglect of study throughout the term. This is not due to the fact that final examinations are given but that undue emphasis is placed upon them and that the teacher has failed to check up on the student's work day by day throughout the term.

One of the points which may be made in favor of final examinations is that they furnish an immediate incentive for review and organization of the content of the course. The writing of an examination itself may be an important part of the student's learning. This is particularly true in the case of questions which require reasoning and organization of information. "There is no impression without expression," and the writing of a three-hour examination is undoubtedly an intensive form of expression. Hence, one is justified in maintaining that written examinations tend more to encourage desirable mental processes than undesirable ones.

III. Passing the final examination an undesirable objective.

The assertion has been made that when a final examination is required, the passing of it tends to become the objective for which many students work. When this occurs it is due not to the fact that the final examination is required but rather to the undue emphasis which is placed upon it by the school. If an examination consists of appropriate questions it is not undesirable to have the student keep it in mind as one of the objectives to be attained by studying the subject-matter of the course. However, as we shall show later, (see page 25) the usual practise is to count the final examination grade as one-third in determining a student's final standing. In many schools it receives less weight. When the final examination counts only one-third or less in determining a student's final standing it is difficult to say in what respect it forms an important educational objective.

IV. Examinations injurious to health of students.

Some critics claim that written examinations, particularly those given at the end of a course, are injurious to the health of students, many of whom make very strenuous preparation for them. The obvious strain which accompanies the writing of answers to the questions of examinations sometimes lasting two or three hours

must also be borne. It is undoubtedly true that both the preparation and the writing frequently make enormous drains on the energies of students. However, no careful investigation has been conducted of the actual effect upon their health. To one who observes the great expenditures of time and energy devoted to social and athletic activities, it is difficult to believe that examinations are in general more injurious to the health of students than many other activities in which they are permitted and even encouraged to engage. Here again it should be realized that this criticism is not fundamentally a criticism of examinations, but rather of setting very long examinations or of placing extreme emphasis upon them by making the final grade of the course depend wholly or very largely upon the examination grade.

V. Time devoted by teachers to written examinations not profitably spent. In the opinion of some critics the time given to the preparation of questions and particularly to the marking of examination papers might be more profitably employed. Information concerning the time actually devoted to the preparation and the administration of written examinations is given in Chapter III. However, it may be pointed out here that a teacher can not attain a high degree of efficiency as an instructor unless he checks up the work of his students in order to assist those who need supplementary and remedial instruction. Only by knowing the extent to which his students have achieved individually and collectively can a teacher make his instruction fit the needs of his class. Thus considerable time must be given to measuring the results of teaching. This is an indispensable portion of the teacher's task. It is only when a teacher devotes an undue proportion of his time to the preparation and administration of examinations that such work tends to be wasted. Doubtless, the time devoted to written examinations might in many cases be profitably increased. Students receiving low marks should have their answers studied in order to ascertain in what ways and why they have failed. Such information will frequently be exceedingly illuminating to the instructor, and aid him in determining his own shortcomings.

CHAPTER III

PREPARATION AND ADMINISTRATION OF EXAMINATIONS IN HIGH SCHOOLS

The data collected. The purpose of the study reported in this chapter was to secure information concerning the present practise in the preparation and administration of written examinations in high schools. A questionnaire was mailed in the fall of 1922 to 254 high-school principals in Illinois and a second one was sent to approximately 2900 high-school teachers.¹ One hundred and eighty-nine replies were received from principals and 1816 from teachers. Of the latter it was necessary to discard eighty so that the following report is based upon returns from only 1736 high-school teachers who are distributed as follows:

Commercial Subjects.....	192	Modern Languages.....	82
Drawing and Art.....	26	Music.....	21
English.....	342	Science.....	309
Home Economics.....	143	Shop Work.....	58
Latin.....	118	Social Science.....	198
Mathematics.....	247		

Representative character of data collected. The high schools from which answers to the questionnaire were received ranged from those established in rural communities to a large metropolitan high school. No supplementary investigation was made to ascertain the extent to which the replies were representative of conditions in Illinois but in the tabulations there was no indication that the data collected were not representative of the state. A few of the replies, particularly those of teachers, suggest that some slight misinterpretation of certain of the questions may have been made. (See page 23). Such cases, however, were relatively rare and probably did not affect the median of the results.

Extent of the requirement of final examinations in Illinois high schools. Evidence of the subjectivity of the marking of examination papers, together with other adverse criticisms of written examinations, has tended to cause many teachers and superintendents to be skeptical of their value. In a number of schools

¹These questionnaires are reproduced in the appendix on pages 66 and 68.

final examinations have been abolished or made optional with the teachers and they are not considered essential by many teachers. In order to ascertain the present practise in Illinois the high-school principals were asked, "Do you require your teachers to give final examinations?" Only twenty-one principals or 11 percent stated that final examinations were not required. Thus it is the practise in Illinois high schools to require that final examinations be given. This, however, does not mean that all students must take them. Of the 168 high schools in which final examinations are required 101 or 60 percent reported that it was their practise to exempt certain students. Scholarship, that is making a grade on daily work above a certain average, was mentioned by all of these schools as one of the conditions on which exemption was based. Deportment was mentioned by 52 percent and attendance by 32 percent as additional conditions.

No information was secured with reference to the explanation of the exemption from examinations of students meeting certain conditions but general observation has indicated that two reasons are frequently recognized. The first is that promise of exemption from the final examinations operates as a powerful motive to secure a high quality of daily work, regular attendance, and good deportment. The other is the belief held by many teachers that final examinations are unnecessary to determine a student's standing in a course. They contend that the average of a student's daily grades should be taken as a final grade for the course.

There is no doubt that the promise of exemption from the final examination operates as a powerful motive in the case of many students. It should, however, be recognized that such an incentive is artificial and therefore open to criticism. In so far as possible a student should be actuated by motives which sustain an intrinsic relation to the subject-matter. If it is necessary or advisable that the final examination be considered as a motive, it could be used to encourage systematic review and organization of the course. This should constitute a very important phase of studying. Students may, of course, be asked by their teachers to review frequently and to summarize and organize the work at the end of the term, but they cannot be convinced easily of the necessity of such work if it receives no more weight in determining their final grade than their performances during an equal period of time elsewhere in the course.

The second reason is a valid one in many cases. In the experience of most teachers the mark made on the final examination changes the standing of relatively few students. Experienced teachers can under favorable conditions estimate with considerable accuracy the achievements of their pupils. If the class is reasonably small and if the teacher has used methods of instruction which call for frequent oral and written performances by the students and has kept a careful record of these performances throughout the term, his estimates will generally be relatively accurate measures of the achievements of the students. There are, however, certain limitations which should be noted. Teachers may be unduly influenced in their estimates by the more recent performances of their students. Unless careful records have been kept throughout the term inferior work at the beginning tends to be overshadowed by good or excellent work during the closing weeks. In case the class is a large one the teacher may not have an adequate opportunity for becoming acquainted with all of its members.

Teachers' estimates are likely to be materially affected by personal characteristics of students; one with a pleasing personality is in many cases rated higher than one who is unattractive. If the classwork is conducted so that there is little or no written performance required, teachers' estimates will necessarily be based almost wholly on the oral responses given during the class period. Some pupils make a good showing in class when the recitation is oral but are at a decided disadvantage when asked to record their answers in writing. Frequently this difficulty is encountered when they are careless in their thinking and do not have clear ideas to express. In oral recitation they are able to make a fair showing because of personal characteristics and because of the stimulus of detailed questioning by the instructor. Furthermore, in a class discussion a bright student who has a good command of language may easily pick up ideas from other members of the class and recall ideas from his general experience sufficient to make a good showing. On the other hand there are students who express themselves more effectively in writing. They may be good thinkers but a little slow in their mental processes and not clever in discussion. Thus there are cases in which it is difficult or impossible for a teacher to estimate accurately the real achievements of students from their daily work alone. The final examination at the

end of the term will in a considerable number of cases furnish additional information which is needed in arriving at the student's true standing.

The final examination in itself provides a distinct type of educational opportunity which does not occur elsewhere in the course. Altho the writers have no evidence to present upon this point they are convinced from their experience with college students and from the comments of a number who have been exempted from final examinations in high school that the practise deprives students of an important educational opportunity. Not infrequently students who have been "excused from examinations" in high school state that they experienced a distinct handicap when they entered college. If final examinations can be justified they should be required of all students. To use them only as a device for motivating the work of the term destroys much of their value.

Time devoted to written examinations. Three questions were asked relative to the time devoted to the preparation and administration of written examinations. The replies from the principals indicated that the most frequent practise is to allow ninety minutes for the writing of a final examination. This is the time allowed in 45 percent of the schools having final examinations. Fifteen percent allow eighty minutes and a slightly larger percent one hundred and twenty minutes.

The teachers were asked to state approximately the number of minutes which they use "in preparing questions for a final examination which students are allowed a total of ninety minutes to answer." The median time which varies only slightly for the different subjects is approximately fifty minutes. Individual teachers in the same subject differ widely in the amount of time which they give to this phase of their work. Two teachers, one in mathematics and one in science, stated that they spent more than six hours in the preparation of a set of final examination questions. In each subject there were a number of other teachers who stated that they devoted not more than thirty minutes to such work. It is possible that some teachers failed to interpret this question correctly but doubtless much of the variation is due to differences in the practises of teachers during the semester. Some probably make a memorandum of questions as they occur during the term and use this list as a basis for preparing the final examination. Also experience is a contributing factor. Teachers who have be-

come very familiar with the subject should be able to formulate questions more quickly than those who are not so well versed.

The teachers were asked also to give the approximate time which they used in "marking the papers of a final examination which students are allowed a total of ninety minutes for answering." They were directed to base their estimates upon a class of twenty-five students. The median time is approximately two and one-half hours. The variations between the different subjects are not large when the differences in their character are considered. The greatest number of hours are required for English and social science and the least for drawing and art. Here also there were wide variations in the amount of time reported by the individual teachers. A total of twenty-five teachers in which all subjects except home economics, shop work, and social science were included, stated that they devoted not more than thirty minutes to marking a set of papers for twenty-five students. On the other hand, thirty-nine teachers stated that they spent 480 minutes or eight hours in the marking of a single set of papers.

It is obvious from the replies received that some teachers treat this phase of their work much more seriously than others or that they employ widely different methods. Probably some correct all errors or insert references which will enable the students to correct their own errors when the papers are returned. Others merely check the errors and still others probably do not attempt to even check each error but estimate the worth of the paper as a whole. The question concerning the amount of time which a teacher is justified in devoting to the marking of a set of examination papers may very profitably be raised. Final examination papers should be treated seriously and there should be an earnest endeavor on the part of the teacher to estimate as accurately as possible the grades which are assigned but it is doubtful if the expenditure of as much as twenty minutes per paper which was reported in some cases could be justified. The median practise seems to represent a more reasonable amount of time.

Characteristics noted in marking examination papers. The principals were asked to state whether it was the practise in their schools for teachers to subtract from a pupil's grade for (1) poor writing, (2) poor spelling, (3) poor English. Seventy-one principals or 42 percent stated that teachers were accustomed to make deductions for poor writing. In 60 percent of the schools it was the

TABLE I. PERCENT OF TEACHERS REPORTING INTENTIONALLY
LOWERING A STUDENT'S GRADE FOR POOR WRITING, POOR
SPELLING, AND POOR ENGLISH

Subject	Poor Writing	Poor Spelling	Poor English
Ancient Language.....	31	75	80
Commercial Subjects.....	61	74	72
Drawing and Art.....	65	82	67
English.....	60	96	98
Home Economics.....	37	72	73
Mathematics.....	31	48	48
Modern Language.....	31	82	74
Music.....	48	60	48
Science.....	35	53	56
Shop and Vocational.....	32	77	74
Social Science.....	37	61	65

practise to lower a student's grade for misspelled words and in 68 percent for poor English. Fifteen principals or 9 percent reported that all three characteristics were recognized only in the marking of papers in courses in English.

The teachers were asked if they intentionally lowered a student's grade because of each of the three characteristics mentioned in the above paragraph. A summary of their replies is given in Table I, which indicates considerable variation with reference to the influence of poor writing, poor spelling, and poor English upon examination grades. Since writing, spelling and English may be considered essential parts of courses in English we should naturally expect that teachers of this subject would intentionally lower a student's grade for defects in any of these characteristics. Outside of the subject of English, the majority of teachers do not lower a grade for poor writing except in commercial subjects, and drawing and art. With the exception of mathematics deduction is made by most teachers for poor spelling. The potency of poor English in determining a student's grade is slightly less than that of spelling in a number of subjects.

The handwriting, spelling, and quality of English which a student uses in writing an examination should be recognized. It does not, however, follow that a student's standing should be intentionally lowered for poor handwriting, poor spelling, and poor English in school subjects other than English. When this is done his grade becomes a measure of these abilities as well as of the abilities in the field of the subject in which the examination is given. In history, for example, a student's grade would become a

composite measure of his achievement in history, the legibility of his handwriting, the quality of his spelling and the use of grammatically correct English. As a result both the teacher and the student are likely to be confused concerning the shortcomings of the examination paper. A better procedure would be to keep a record of the errors in spelling, poor writing, and poor English and when it is considered desirable a separate grade may be given covering these three characteristics. Credit for a course may be withheld until the student has brought his handwriting, spelling and English up to a satisfactory standing.

The weighting of questions. Sixty-four percent of the high school principals indicated that their teachers were accustomed to give more credit for correct answers to difficult questions than to easy ones. Approximately four-fifths of the teachers replying to the questionnaire stated that they attempted to weight examination questions on the basis of difficulty. Thus there is a very definite effort to eliminate the errors introduced in examination grades by the unequal difficulty of questions. (See page 10.)

Recognition of rate of work. Eighty-two percent of the teachers stated that they were accustomed to set examinations short enough so that practically all students could answer all the questions. Only 32 percent noted the time which each student spent in writing his examination paper, and only 8 percent said it was their custom to set examinations long enough so that practically no student would have time to answer all of the questions. Thus it is clear that relatively few teachers recognize the student's rate of work in determining his grade on an examination.

Incidentally it may be noted that when examinations are short enough so that practically all students can finish a great deal of time is wasted. Individual differences exist in all classes and it is not at all unusual to find some student finishing in one-third to one-half of the time which others devote to the examination. Aside from the waste of time which results from this practise, it is likely that the confusion caused by the leaving of those pupils who have finished tends to disturb the attention of those who are still writing. If final examinations constitute a valuable educational opportunity, there is no justification for wasting time. It is much better to set an examination long enough so that practically all students will be occupied for the entire period.

TABLE II. PERCENT OF TEACHERS GIVING AFFIRMATIVE ANSWERS
TO FOUR QUESTIONS RELATING TO THE MARKING OF EXAMI-
NATION PAPERS

Subject	Correct Answers Written	Each Ques. on one paper	One Ques. on all papers	Each paper as a whole
Ancient Language.....	9	73	24	20
Commercial Subjects.....	48	82	16	31
Drawing and Art.....	40	75	25	42
English.....	22	76	19	34
Home Economics.....	19	68	24	38
Mathematics.....	72	75	18	24
Modern Language.....	15	74	23	33
Music.....	35	80	21	42
Science.....	27	72	27	27
Shop and Vocational.....	36	75	27	37
Social Science.....	18	75	22	40

Method of marking examination papers. Scientific investi-
gation has revealed that the reliability of examination grades can
be materially increased by adopting a systematic method in mark-
ing papers.² Among the procedures recommended are the writing
out of correct answers, and the grading of one question on all of the
papers before taking up another question. In order to ascertain
the practise of high-school teachers in marking examination papers
the following four questions were included in the questionnaire
sent to them.

1. Before starting to grade a set of examination papers do you
write out the answers which you consider correct?
2. Do you usually grade all the answers on one paper before
taking up those of another paper?
3. Do you usually grade the answers to one question on all
of the papers before taking up the answers to a second
question?
4. Instead of marking the answers to each question separate-
ly do you attempt to estimate the value of the paper as
a whole?

The percent of teachers giving affirmative answers to these ques-
tions is given in Table II. A few apparent discrepancies in this
table are due to the fact that certain teachers did not answer
all the questions. Normally we should expect a teacher who an-
swered the second question affirmatively to answer the third one
negatively. This, however, did not always happen.

²Kelly, F. J. "Teachers' marks," Teachers College Contributions to Education,
No. 66. New York: Teachers College, 1914, 83 p.

With the exception of mathematics, it is not the custom of teachers to write out the answers to their questions. No data are at hand to show what effect this has upon the accuracy of the examination marks. Experience with standardized tests would indicate that the failure to write out correct answers, at least in abbreviated form, would operate to make the grading of examination papers less accurate.

About three-fourths of the teachers are accustomed to mark all the questions on one paper before taking up another. This plan has the advantage of enabling the teacher to consider a pupil's performance as a whole. In the case of students who make a large number of errors the teacher will find this helpful in providing remedial instruction. It has, however, been proposed that the reliability of examination grades can be increased by marking the answers to one question on all the papers before taking up the answers to another question.

Directions to students concerning methods of work. The high-school teachers were asked the following question, "Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answer, arrangement of work, etc.)" Only 35 percent of the teachers gave an affirmative answer. It is possible that in some classes there is a sufficiently definite understanding concerning the methods of work to be followed and explicit directions are unnecessary. However, it is likely that in many cases students would be able to give more truthful evidence of their ability if they were given precise directions concerning the length of answer, desired arrangement of work, etc. The order in which the questions are to be answered is a point which should be stressed. In the case of questions that are at all indefinite or general there should be specifications concerning the degree of elaborateness which is expected in the answer.

Recognition of a standard distribution in assigning grades to examination papers. The teachers were asked the following question, "In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?" Only 31 percent of the teachers gave an affirmative answer to this question. This probably means that

relatively few teachers have recognized the distinction between "scores" and "grades," (See pages 11-13 for an explanation of this distinction.) and for this reason are neglecting one means of making their grades more accurate measures of school achievement.

Relation of examination grades to final grades. The principals were asked if they advised their teachers as to the proportion of the final mark for the semester which should be based upon the final examination. Only eleven or 7 percent replied negatively. Of those who gave advice on this matter 95 percent made a definite ruling. The most frequently mentioned proportions allotted the final examination are 25, 30, $33\frac{1}{3}$, and 40 percent. In 4 percent of the schools the examination counts for one-half in determining a student's final grade, in 1.3 percent for only one-tenth of the final grade. The teachers were also asked this same question. The replies varied from 10 percent to 50 percent. Except in science and shop work, the median practise is to estimate the final examination mark as one-third in determining a pupil's final grade. A considerable number of teachers indicated that they gave not more than one-fourth or one-fifth value to the final examination.

Summary. The typical practise with reference to final examinations in Illinois high schools may be summarized as follows:

1. Final examinations are required of students and exemptions are made largely on the basis of scholarship.

2. Students are allowed ninety minutes for writing a final examination. Teachers spend slightly less than one hour in preparing examination questions and from two to three hours in grading a set of papers for twenty-five students.

3. With the exception of mathematics, the majority of teachers lower a student's grade for spelling, and with the exception of mathematics and music the majority lower it for poor English. Poor writing is not a potent factor except in English, commercial subjects, and drawing and art.

4. About three-fourths of the teachers attempt to weight the questions on the basis of difficulty.

5. The majority of the teachers do not consider rate of work in estimating the grade assigned the final examination paper.

6. The majority of teachers do not write out the answers to the questions preparatory to the marking of papers. The general practise is to mark all answers on one paper before taking up the next.

7. About one teacher in three writes out directions with reference to the procedure which the students shall use in answering the questions.

8. About one teacher in three tries to have his grades conform to a standard distribution.

9. The proportion of the final mark for the semester which is based upon the final examination grade varies from 10 to 50 percent. The median practise is $33\frac{1}{3}$ percent. The majority of the principals make a definite ruling regarding the value placed upon the final examination.

CHAPTER IV

THE CONSTANT AND VARIABLE ERRORS IN EXAMINATION GRADES

Constant and variable errors of measurement.¹ Two types of errors are encountered in educational measurement. The presence of *variable errors* is indicated when a test is given twice to the same group of pupils. The two average scores of the group may be the same but this will not be true for individual pupils. A few pupils will make the same or approximately the same score on the two trials. Others will make higher scores on the second trial than on the first, while still others will make lower scores on the second trial than on the first. If we assume that the average of the two scores obtained represents an approximately true measure of a pupil's achievement then the differences between the first set of scores and the corresponding average scores would be the variable errors of the measures resulting from the first application of the test. Some of these differences approximate zero. Some of them are positive and about an equal number negative. Another set of variable errors would be obtained by using the scores secured from the second application of the test. In the case of a number of pupils the variable errors for the first application of the test will not be the same as those for the second application. Thus, as the name implies, variable errors change in magnitude from pupil to pupil within a group and also for the same pupil in a series of measurements of the same achievements.

A *constant error* is the same for all members of a group. Such an error occurs in teachers' marks where there is a tendency to grade too high or too low. It is found in the case of standardized educational tests when mistakes occur in the time allowed or when other departures are made from standard testing conditions. A constant error may be either positive or negative and it is generally different for different tests.

¹For a more detailed discussion of the nature and magnitude of the constant and variable errors of educational measurement, see Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923. 30 p.

These two types of errors usually occur in combination, that is, a given measurement may and frequently does involve both a constant error and a variable error. The actual error is a combination of these two. However, in studying the accuracy of educational measurements it is helpful to distinguish between the two types and to consider each separately. The usual method used for calculating an index of the magnitude of the variable errors does not give any indication of the magnitude of the constant error. Also the method commonly used for determining the presence and probable magnitude of constant errors does not yield an index of the variable errors. Furthermore, different methods are required for decreasing the two types of errors in educational measurements.

Methods of describing the magnitude of the variable errors of measurement yielded by standardized educational tests. In describing the magnitude of variable errors in the measures yielded by standardized educational tests, the usual method is to have the test given twice to a typical group of pupils under as nearly the same conditions as possible. The coefficient of correlation between the two sets of measures is taken as the index of the magnitude of the variable errors. Usually there will be a constant error in one and sometimes in both of the sets of measures but the statistical procedure used is such that this error does not affect in any way the coefficient of correlation secured. This coefficient of correlation is commonly spoken of as the coefficient of reliability. A coefficient of 1.00 would mean that the variable errors were zero.

Available data with reference to the magnitude of errors of examination grades and standardized test scores not comparable. Investigations of the Starch-Elliott type have proven that examination grades involve errors but the method which they employed is different from that used in studying the errors in the scores yielded by standardized educational tests. Starch and Elliott confined their efforts to a study of the subjectivity of the marking of a single examination paper. Except in the use of quality scales, as in handwriting and English Composition, the scoring of standardized educational tests has been made highly objective. Hence, there has been little need for studying the subjectivity of the marking of test papers. On the other hand, there

has not been, so far as the writers are aware, any reported attempt to apply to written examinations the method commonly used in studying the reliability of standardized educational tests. Hence, comparisons to show the relative reliability of the two types of measuring instruments cannot as yet be made.

For this reason in the present investigation it has seemed worth while to apply to written examinations the same method which is commonly used in studying the reliability of educational tests. Certain modifications are of course necessary. These will be noted in the following paragraphs. The investigation pertains primarily to the variable errors involved in examination grades. Incidentally some light will be thrown upon the magnitude of constant errors.

Methods employed in the present investigation of the reliability of written examinations. The essential feature of the methods employed in the present investigation is securing two independent examination grades for each pupil for the same units of work. This requires that two examinations be given to each of the groups of students from which data were secured. Two methods were used. These are described in the following directions which were sent to those cooperating in this investigation.

METHOD I

Two sets of examination questions are to be prepared by a single person, or two or more persons working together. Each of the two lists should contain the same number of questions. There should be a distinct effort to make the two lists approximately equal in difficulty and as nearly as possible similar in respect to the type of questions.

After the two lists of questions have been made both should be given by each teacher to all of her pupils under as nearly the same conditions as possible. If not given on the same day, the two examinations should be given within a period of one week. For example, if two sets of examination questions in seventh grade geography have been prepared, both sets of questions should be given by each seventh grade teacher to all of her pupils.

Each teacher is to mark both sets of examination papers for her pupils. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination paper. When two or more teachers have given the same examinations it is not necessary that they confer in regard to the marking of the papers. If this is done a memorandum regarding the procedure should be attached to the examination papers.

This method may be followed by a single teacher who has two or more sections of a given subject. Both examinations should be given to all sections taught by this teacher. This method of studying the reliability of written examinations can be applied to any school subject. The Bureau of Educational Research is most interested in having it applied to arithmetic, history, geography, and language in the elementary school and to history, mathematics, English, and science in the high school.

METHOD II

Two sets of examination questions for the same subject are to be prepared by two teachers working independently, each teacher preparing a set. There is no requirement concerning the length or the difficulty of the two sets of examination questions except that both should cover the same amount of work. The teachers who prepare the questions should not confer concerning either the kind or the number of questions asked.

After the questions have been prepared, both sets are to be given by each teacher to all of her pupils. If not given on the same day, the two examinations should be given within a period of one week.

After the examinations have been given each teacher will grade all of the papers written upon the questions that she prepared. This will mean that she will grade a set of papers for her own pupils and also a set for the pupils of the other teacher. There should be no conferring between the teachers in regard to the method of scoring. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination papers.

The data collected. Through the city superintendents and high-school principals a general invitation was extended to school systems in Illinois to participate in this investigation at the close of the second semester 1921-22 and also at the close of the first semester 1922-23. No instructions other than those just noted were given to those who cooperated. It should, therefore, be borne in mind that the data collected are for written examinations as they are usually given and not for special types of examinations or for unusual methods in the administration or the grading of the test papers. The reliability of examination grades could probably have been increased if certain directions had been formulated in regard to the marking of the examination papers but the purpose of the investigation was to determine the reliability of typical written examinations administered in the usual way.

Returns were secured from seventy-two groups of children but it was necessary to discard the data for six groups because

instructions had not been followed. The examinations given to the sixty-six groups were all of the traditional type. The papers were marked by the teachers on the scale of 100 percent, and were then sent to the Bureau of Educational Research. The coefficients of reliability reported in this chapter were calculated under the direction of the writers.

Coefficients of reliability of examination grades. The coefficients of reliability of written examinations for the sixty-six groups of students are summarized in Table III. This table also shows the number of students in each class, the number of questions in each examination and the method followed in giving the examination. The reliability coefficients have been grouped by subjects and have been arranged in descending order of magnitude. For those entries marked with an asterisk (*) in the column headed "Method," one of the examinations was given by the principal or some other person not actually teaching a class in the subject at that time. However, this person was considered competent to be in charge of the examination. The total distribution of reliability coefficients is given in Table IV.

Two of the coefficients of reliability are negative. The highest is .95. It is interesting to note that the coefficients given for history are, on the average, higher than those obtained for arithmetic. The most reliable examinations given were in algebra. With the exception of history, arithmetic, and algebra, the number of groups is so small that comparisons can not have much significance. The median coefficient of reliability .65 may be used as a general index of the reliability of written examinations.

The coefficients of reliability of standardized educational tests. McCall² has stated that the "range of self-correlation for many standardized tests is about .5 to about .9." The writer's experience has indicated a somewhat greater range. In Table V the reliability of a number of standardized educational tests is given. Those for the silent reading tests by Brown, Starch and Courtis are taken from a recent bulletin³ by the writer. The range in this table is from .19 to .92.

²McCall, W. A. *How to Measure in Education*. New York: The Macmillan Company, 1922, p. 396.

³Monroe, Walter S. "A critical study of certain silent reading tests." *University of Illinois Bulletin*, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p.33-34.

TABLE III. COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
64 (Arithmetic).....	21	10	.76	II
58 ".....	35	7 and 8	.74	II
43 ".....	24	10	.73	II
63 ".....	38	10	.71	II
61 ".....	64	10	.69	II
2 ".....	37	5	.67	I
62 ".....	38	10	.64	I
1 ".....	88	10	.64	I
60 ".....	41	5 and 10	.61	II
65 ".....	33	10	.60	II
4 ".....	72	6	.56	I
71 ".....	17	10	.48	*
44 ".....	22	7 and 10	.48	II
3 ".....	55	6	.47	I
66 ".....	73	10	.47	II
23 ".....	27	5	.33	I
32 ".....	56	5	.30	II
59 ".....	21	6 and 10	.29	I
45 ".....	43	10	.06	II
42 ".....	27	8 and 10	— .18	II
5 (Algebra).....	51	10	.92	I
72 ".....	20	30 and 19	.91	*
50 ".....	23	5 and 7	.88	II
34 ".....	37	10 and 6	.82	II
57 ".....	36	10	.81	II
6 ".....	52	6	.78	—
35 ".....	45	15	.73	II
31 ".....	54	6	.61	II
7 (Language).....	33	5	.68	—
9 ".....	54	5	.62	—
8 ".....	53	5	.57	—
10 (Literature).....	74	5	.19	I
18 (English).....	15	3	.78	I
12 ".....	43	5	.53	II
11 ".....	27	5	.53	I
54 ".....	37	7 and 45	.50	II
53 ".....	21	6 and 7	.47	II
13 ".....	52	5 and 8	.41	II

TABLE III. (CONTINUED) COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
20 (History)	14	5	.95	—
15 “	28	5	.85	I
23 “	19	5	.78	I
36 “	43	5	.76	II
56 “	19	9	.75	*
67 “	53	5	.75	II
41 “	32	8 and 10	.67	II
40 “	28	5 and 7	.66	II
14 “	64	5	.63	I
48 “	24	10	.57	II
49 “	30	5 and 10	.55	II
39 (Geography).....	29	10	.66	II
46 “	29	5 and 10	.66	II
47 “	47	5 and 10	.62	II
16 “	21	5	.43	I
68 “	23	10 and 8	.32	*
38 “	26	10 and 5	— .11	II
17 (Civics)	63	5	.39	I
37 (Latin)	31	8	.89	II
19 “	30	2	.82	I
51 “	50	6 and 8	.68	II
21 (Spelling)	33	5	.87	I
52 (Geometry).....	42	5	.75	II
70 “	10	8	.68	*
22 “	46	4 and 5	.34	—
24 (Spanish)	18	5	.79	I
33 (German)	23	7 and 8	.83	I
69 (Commerce).....	16	16 and 10	.53	*

TABLE IV. SUMMARY DISTRIBUTION OF COEFFICIENTS OF RELIABILITY FOR WRITTEN EXAMINATIONS

Size of Coefficient of Correlation	Frequency
.95	1
.90	2
.85	4
.80	4
.75	9
.70	4
.65	9
.60	8
.55	4
.50	4
.45	5
.40	2
.35	1
.30	4
.25	1
.20	0
.15	1
.10	0
.5	1
0	0
— .5	0
— .10	0
— .15	1
— .20	1
Total.....	.66
Median.....	.65

From certain unpublished studies by the writer the following information has been obtained. The Courtis Standard Research Test, Series B, Forms 1 and 2 were given to pupils as follows: Grade V, 89; Grade VI, 81; Grade VII, 52; and Grade VIII, 38. The thirty-two coefficients of reliability ranged from .409 to .904 with the median at .665. Forms 1 and 3 were given to a slightly larger group in each of the four grades. The thirty-two coefficients of correlation between the two sets of scores for this administration of Series B ranged from .528 to .963 with the median at .704. The Woody Arithmetic Scales, Series A, were given to several groups of pupils. Two scores were secured by using alternate items of each of the scales and applying Brown's formula.⁴ The twelve coefficients of reliability computed in this way ranged from .91 to .46 with the average at .66. Forms 1 and

⁴ $r_{12} = \frac{2r_h}{1+r_h}$ In this formula r_h is the correlation between two scores which this test yields. One is based upon reproduction and the other upon answers to questions.

TABLE V. RELIABILITY COEFFICIENTS OF STANDARDIZED EDUCATIONAL TESTS

Test	Coefficient
Illinois General Intelligence Scale*	.92
Courtis Standard Research Tests, Series B†	.87
Brown Silent Reading Test—Rate	.86
Courtis Silent Reading Test, No. 2—Rate	.85
Otis Group Intelligence Scale§	.84
Monroe Standardized Silent Reading Test Revised*—Rate	.84
Courtis Silent Reading Test, No. 2—Comprehension—No. Quest	.80
Starch Silent Reading Test—Comprehension—Words	.77
Monroe General Survey Scale in Arithmetic*	.76
Monroe Standardized Silent Reading Test Revised*—Comprehension	.76
Monroe Standardized Silent Reading Test Revised*—Rate	.75
Monroe Standardized Silent Reading Test Revised*—Comprehension	.72
Starch Silent Reading Test—Comprehension—Ideas	.72
Indiana Attainment Scale, No. 1†	.66
Starch Silent Reading Test—Rate	.62
Pressey Primer Scale†	.59
Courtis Silent Reading Test, No. 2—Comprehension—Index	.58
Pressey First Grade Vocabulary Scale†	.37
Brown Silent Reading Test—Comprehension—Quantity	.36
Pressey Primer Scale†	.33
Brown Silent Reading Test—Comprehension—Quality	.19

*Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 47.

†Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades: its validity and reliability," Journal of Educational Research, 1:285-94, April, 1920.

‡Unpublished data of the Bureau of Educational Research, University of Illinois.

§Colvin, S. S. "Some recent results obtained from the Otis Group Intelligence Scale," Journal of Educational Research, 3:1-12, January, 1921.

2 of Monroe's Standardized Reasoning Test in Arithmetic were given to pupils as follows: Grade V, 36; Grade VI, 92; Grade VII, 76; Grade VIII, 81. The coefficients of reliability for correct principle were as follows: .530, .630, .645, and .723; for correct answer they were .518, .528, .576, and .707. Using Brown's formula the coefficients of reliability for Gray's Silent Reading Tests were computed for thirty grade groups. These coefficients ranged from .55 to .85 with the median at .67. The number of pupils per group was less than 100 in only five cases. For several grade groups reliability coefficients were secured for Monroe's Standardized Silent Reading Tests which ranged from .222 to .907 with an average of .669.

Haggerty has computed the reliability for both Sigma 1 and Sigma 3 of his Reading Examination by having the same test repeated. In the case of Sigma 1 the interval between the two applications of the test was six weeks. For 200 children in Grades I to III inclusive the coefficient of reliability .84 was obtained. In the

case of Sigma 3 the interval between the two applications was only two days. For 126 pupils from Grades V to VIII, inclusive, the coefficient of reliability was found to be .885. For the sentence test alone the reliability coefficient was .769 and for the paragraph test, .806. For Thorndike's Scale Alpha for the Understanding of Sentences, McCall has reported a coefficient of reliability of .37. This was obtained by using a test similar to Alpha but not considered a duplicate form. Gates⁵ reported reliability coefficients for Thorndike-McCall Reading Scale which ranged from .25 to .72. All of these were for pupils belonging to a single grade. For the Burgess Picture Supplement Scale the author has given coefficients of reliability ranging from .62 to .88 for grade groups from the second to sixth grades inclusive. In each case the number of pupils was relatively small. Gates gave coefficients of .62, .59 and .66 for three grade groups.

For the Otis Self-Administering Test of Mental Ability the author has reported an average reliability coefficient of .921 for the higher examination and of .948 for the intermediate examination. Presumably these coefficients are based on the scores secured from pupils for a sequence of several grades. For the separate tests of the Stanford Achievement Test the authors reported coefficients of reliability based upon separate grade groups which ranged from .75 to .96. When the composite score of all the tests was used the reliability coefficient was .98.

The relative reliability of written examinations and standardized educational tests. The data which have just been submitted indicate that the difference between the reliability of the two types of instruments is not as great as is commonly believed. The median of the reliability coefficient for written examinations given in Table IV is .65. There are many reliability coefficients for standardized tests in Table V which are less than this. Furthermore, the additional citations of coefficients of correlation in the above paragraphs indicate that for a number of standardized educational tests which have been very widely used the median of the reliability coefficients for grade groups is in the neighborhood of .65. Thus the conclusion seems justified that altho some of our more elaborate standardized tests, such as the Stanford Achieve-

⁵Gates, Arthur I. "An experimental statistical study of reading tests," *Journal of Educational Psychology*, 12:379, October, 1921.

ment Test, the Illinois General Intelligence Scale, and the Otis Self-Administering Test of Mental Ability, may be expected to yield measures whose reliability is greatly in excess of that of typical written examinations, many widely used standardized educational tests yield measures which possess about the same degree of reliability as the grades obtained from written examinations prepared by teachers and other school officials. It should be noted that reliability refers only to the variable errors of measurement. The constant errors as we shall show, (p. 40) are likely to be very much larger in examination grades than in the scores yielded by standardized educational tests. It should also be noted that the time required to give many of the standardized tests is much less than that devoted to a typical written examination.

The absolute reliability of examination grades. The statement that the reliability of a typical examination is equivalent to that of many standardized tests and only slightly less than that of a number of others still leaves a doubt with reference to the absolute reliability. For practical purposes the reliability coefficient of .65 needs to be interpreted in terms of the variable errors of measurement to be expected. The correlation tables for eight groups having a reliability coefficient of approximately .65 were taken and the scores translated into a five point system of school grades. It is assumed that these classes were typical and the highest scores were translated into a mark of "A," the lowest into a mark of "E." This was done in an arbitrary way but the results indicate roughly one meaning which may be attached to a reliability coefficient of .65. It was found that in 40 percent of the cases the students received the same grade in the two examinations. In an additional 42 percent the grade which they received on the first examination was only one point higher or lower than that received on the second. For example, if a student in this group made a "D" on one examination, he made an "E" or "C" on the other. The two grades received by the remaining 18 percent differed by two points or more.

Conditions tending to produce variable errors of measurement in examination grades. Several sets of examination papers were examined in order to ascertain the conditions which tended to produce the lowest coefficients of reliability and hence the largest variable errors of measurement. The most potent cause

appeared to be that the two teachers recognized widely different educational objectives in making out the two sets of examination questions. This seemed to be the case in Group 42, arithmetic, for which the coefficient of correlation was $-.18$. In Group 22, geometry, there was a difference in the general plan of the examinations; one teacher permitted the students to choose one of two questions in part of the examination while the other required that all questions be answered. This difference in the plan of the examination appeared to increase the variable errors of measurement. There was also a difference in the educational objectives recognized in that one teacher placed much more emphasis upon the practical application of geometry than the other.

Another cause which operated to lower the degree of correlation and hence to increase the magnitude of the variable error was the adherence to different standards of excellence by the teachers who graded the papers. For example, in Group 45, arithmetic, one teacher considered only the final answer to the exercise; if that was right the student received full credit—if wrong, no credit was given. The other teacher gave credit for correct principle. The coefficient of reliability for this group was $.06$.

It was noticed that in general pupils made higher grades on the tests set by their own teacher than on those set by another person. This appeared to be true even when distinct differences could not be identified either in the educational objectives or in the methods of grading of the two teachers. In Group 32 for which a reliability coefficient of $.30$ was obtained when the grades made on the first examination were correlated with those made on the second examination, a second coefficient was calculated by comparing the student's grade made on the examination prepared by his own teacher with that set by another teacher. This procedure gave a coefficient of correlation of $.40$. When the two classes were taken separately coefficients of $.57$ and $.44$ were obtained. These data tend to supplement the evidence already cited that differences in the content of the examination and in the plan of marking are potent factors in producing the variable errors of measurement.

The magnitude of constant errors in examination grades. It is probable that most of the teachers marking the examination papers did not recognize the distinction between "scores" and "grades"⁶ and that the marks placed upon the papers were con-

⁶See page 11 for a statement of this distinction.

TABLE VI. DISTRIBUTION OF DIFFERENCES BETWEEN AVERAGES
OF EXAMINATION GRADES

Difference	Frequency
50	1
32	1
29	1
27	1
22	1
21	1
20	2
18	3
16	1
15	1
14	2
13	2
12	1
10	2
9	1
8	4
7	3
6	6
5	4
4	6
3	5
2	6
1	8
0	3
Total.....	66
Median.....	6.2

sidered as "grades." In several instances the "grades" made on one examination were on the average much higher than those made on the other. If "scores" were used as "grades," any differences between the averages of the two sets of measures indicate the presence of constant errors. In order to secure an index of their magnitude the differences were calculated for the sixty-six groups to which two examinations were given. These are assembled in Table VI. For three of these groups the difference between the averages of the two sets of "grades" was zero; for eight other groups it was one. At the other extreme we find a difference of 50 in the case of one group. The median difference is 6.2.

It should be noted that the differences between the averages of two sets of examination grades are not constant errors. They are merely indicative of the presence of constant errors. If one examination was easy and the other hard the difference would be the sum of a positive error and a negative error. If both examinations were hard the difference would be smaller than the constant error in either average. The large differences shown in Table VI

are probably caused by the combination of an easy examination with a difficult one. This was very obviously true in the case of the one difference of 50. Furthermore, in interpreting Table VI it should be remembered that possibly some of the teachers recognized the distinction between "scores" and "grades," and the marks would have been appropriately adjusted before being used as "grades."

So far as it was possible to ascertain from an analysis of the examination papers the large differences are due to two causes—differences in the difficulty of the two sets of examination questions and in the severity of the grading. For example one of the examinations which produced a difference of 40 consisted of seven questions of which the pupils were permitted to answer any five. These questions were relatively easy. In the other examination, there were ten questions and the pupils were required to answer all of them. Very few were able to complete this second examination in the time allowed and the teacher appears to have counted the unfinished exercises as failures. Nine out of twenty-two children in the second group made zero on the examination. In this way a very large constant error was introduced but the coefficient of reliability for this group was .48.

Relative magnitude of constant errors in examination grades and in standardized test scores. In another place⁷ the writer has discussed the magnitude of the constant errors in educational tests. In cases where there has been coaching for tests, intentional or not, or disregard for standard directions, large constant errors have been introduced. In one extreme instance a constant error of over three and a half years occurred in the mental age scores of a group of children. In general, however, because of the standard directions for administering the tests and scoring the papers, of the objectivity of the marking, and of the norms for interpreting test scores, the constant errors in standardized tests are very much smaller, and are likely always to be smaller than those found in examinations given by teachers. However, some reduction in the magnitude of the constant errors in examination scores will result when the use of either very easy or very difficult sets of questions is avoided and when a conservative plan of marking is followed.

⁷Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923, p.19-20.

Explanation of the apparent contradiction between the results of this investigation and previous studies of examination grades. The results of this investigation have caused the writers to revise their estimate of the accuracy of examination grades. The findings indicate that the errors are much less than they appeared to be from evidence resulting from investigations of the Starch-Elliott type. One naturally asks the question, "Why this apparent contradiction?" Starch and Elliott obtained similar results for three different examination papers and numerous other investigators have corroborated their findings. The mass of evidence accumulated is so extensive and uniform in character that one would naturally be inclined to accept the conclusions indicated in preference to the apparent contradictory results of the present investigation. However, a careful analysis of the procedures reveals that the results are not necessarily contradictory. The method followed by Starch and Elliott combines both constant errors and variable errors. The "grades" assigned to the examination paper in geometry were influenced both by the subjectivity of the marking and by the tendency of some teachers to grade high and of others to grade low. The present investigation has separated the variable errors from the constant. It has also shown that the examination scores have in some cases involved relatively large constant errors. The extreme differences between the grades assigned to the same paper reported by Starch and Elliott (see page 9) are easily explained when it is understood that they represent the combination of variable errors and constant errors. Especially is this true when we realize that the constant errors would likely be larger for teachers of different schools as in their investigation than for teachers in the same school as in the present investigation.

Conclusion with reference to relative accuracy of examination grades and scores yielded by standardized tests. As already indicated the writers believe that the data presented in this chapter show that examination grades are more accurate measures of achievement than many persons have considered them to be. Standardized tests yield scores involving errors, both constant and variable, but in the case of our best standardized tests these errors are distinctly less than the corresponding errors in examination grades. Furthermore, measurement by means of standardized tests usually requires much less time than is com-

monly devoted to written examinations. This conclusion refers to written examinations of the traditional type and administered under typical conditions. It is likely that written examinations and their administration may be improved so that the difference in the accuracy of examination grades and test scores will become much less than at present.⁸

⁸The conditions of standardized tests would have been more closely approximated if both sets of examination questions had been prepared by the same person and marked by different persons. If this had been done it is reasonable to expect that the coefficients of reliability would have been somewhat higher and the differences in the averages of the two sets of scores smaller.

CHAPTER V

THE CONTENT OF WRITTEN EXAMINATIONS

The data collected. In response to an invitation sent to superintendents and high-school principals in Illinois sets of examinations were received from fifty-six schools for the first semester and from fifty schools for the second semester of the school year 1921-22. A range of approximately sixty subjects was represented. It seemed desirable to restrict this analysis of sets of questions to the thirteen subjects listed in Table VII. The number of sets of questions and also the total number of questions are given in this table.

Classification of questions according to type. After considerable experimentation a list of fifty types of questions as given below was formulated.

— Aims \ —	How many (tell)
— Analysis	— Law (give the)
— Cause (give)	Mathematical operations of addition, subtraction, multiplication, and division
— Classification	Method
— Comparison	— Outline
— Completion	Parsing
— Conjugation	— Proof
— Construction (a figure, study or statement)	Punctuation (capitalize and correct sentences)
— Construction (give the)	Recall
— Contrast (general)	Reduction to lowest terms
— Contrast (specific basis)	Relationships (give the)
— Correction	Rule
— Criticism	Scanning
— Decision (choice or preference)	Simplification
— Declension	Source
— Definition	Substitution (values for letters)
— Description (characterization)	Summary
— Diagram (illustrate by)	Solving for unknown quantity
— Discussion	Syllabus
— Effect (give the)	Translation (foreign language to English)
— Evaluation	Translation (English to foreign language)
— Example (illustrate by)	Use (give the)
— Expansion	Where (tell)
— Explanation (tell why or how)	
— Facts (definite number)	
— Facts (indefinite number)	
— Factoring	

TABLE VII. NUMBER OF QUESTIONS AND SETS OF QUESTIONS
EXAMINED

Subject	Sets	Questions
English I.....	80	721
English II.....	83	694
English III.....	79	721
Algebra I.....	80	731
Plane Geometry.....	80	636
Latin I.....	81	683
Latin II.....	76	539
Physics.....	76	795
General Science.....	62	789
Civics.....	59	560
American History.....	62	550
Domestic Science.....	42	392
Domestic Art.....	41	368
Total.....	901	7621

All questions for the thirteen subjects mentioned in Table VII were classified under some one of these types. This classification was made by Mr. Souders with the assistance of a single clerk working under his immediate direction. Altho any classification of this kind is necessarily subjective, a relatively high degree of uniformity has, we believe, been secured.

Summary of classification. Twenty-six of the fifty types of questions were represented in six or more of the thirteen subjects. The relative frequency of each is given in Table VIII. This classification of examination questions shows a high frequency of certain types and very little or no use of a number of other types. If we omit Latin, Algebra, and Plane Geometry in which the nature of the subject-matter restricts the kind of question asked, we find that 32 percent of all the questions require "explanation." The next most frequent type used, 21 percent, calls for a "definite number of facts."

Frequently all questions are considered as belonging to one of two groups, "thought questions" or "memory questions." Such a definite classification is not, however, always possible. The character of the mental process involved in answering depends upon the person replying as well as upon the form of the question asked. Those questions calling for definite facts are almost certain to be based upon memory; on the other hand, those requiring classification, evaluation, contrast, etc. are likely to demand

TABLE VIII. PERCENT OF RELATIVE FREQUENCY OF TYPES OF QUESTIONS IN THIRTEEN SCHOOL SUBJECTS

Type	English I	English II	English III	American History	Civics	Latin I	Latin II	Algebra I	Plane Geometry	General Science	Physics	Domestic Science	Domestic Art
1. Analysis.....	1	—	—	—	—	—	—	—	—	2	1	—	1
2. Cause or effect.....	1	—	—	4	—	—	1	—	—	1	—	1	1
3. Classification.....	1	1	1	—	1	—	1	—	—	1	—	2	2
4. Completion.....	2	2	1	—	—	2	5	1	12	2	2	3	10
5. Construction.....	14	13	13	1	1	—	—	—	—	—	—	—	3
6. Contrast (Gen.).....	—	—	—	—	—	2	1	—	—	3	4	—	—
7. Contrast (Spec.).....	2	4	5	3	4	—	—	—	—	—	—	3	—
8. Correction.....	3	3	3	—	—	—	—	—	—	—	—	—	—
9. Decision.....	2	1	—	—	1	—	1	2	5	—	11	1	—
10. Definition.....	5	6	4	1	6	—	2	—	—	4	5	2	2
11. Description.....	6	3	6	2	—	—	—	—	1	6	6	5	6
12. Diagram.....	—	—	—	—	—	—	—	—	—	5	5	1	3
13. Discussion.....	—	—	7	16	7	—	1	—	—	4	6	7	3
14. Effect.....	—	—	1	3	1	—	1	—	—	2	5	2	3
15. Evaluation.....	—	—	—	—	1	—	—	—	—	—	—	—	—
16. Example.....	5	4	3	1	2	3	2	2	1	5	4	—	—
17. Explanation.....	21	22	15	36	38	8	13	1	6	42	37	34	28
18. Facts (Def. no.).....	18	20	20	21	20	9	7	1	4	17	10	25	29
19. Facts (Indef. no.).....	1	1	2	3	2	—	—	—	—	2	1	3	2
20. How many.....	—	—	—	—	—	—	—	—	—	—	—	—	—
21. Outline.....	2	2	4	5	4	—	1	—	—	1	—	2	3
22. Proof.....	—	—	—	—	—	—	—	2	56	—	—	—	—
23. Recall.....	8	7	6	3	2	1	3	2	—	1	1	2	—
24. Rule.....	1	1	—	—	3	2	3	4	2	—	2	1	1
25. Use.....	1	2	1	—	—	3	1	—	—	—	1	2	2
26. Where.....	1	1	—	—	—	—	—	—	1	—	—	—	—
All other types.....	5	3	5	—	1	68	60	86	11	1	12	2	1
Total Questions.....	922	906	802	601	761	760	578	646	699	667	887	500	580
Percent of Memory.....	21	19	16	4	11	73	65	69	74	9	24	9	6
Percent of Thought.....	79	81	84	96	89	27	35	31	26	91	76	91	94

The dash (—) means a frequency of less than $\frac{1}{2}$ of 1 percent.

thought on the part of most students. If, however, such classifications or evaluations have been made in a previous class exercise some students may easily remember the answers and, in such a case, a thought question for one student becomes a memory question for another. For the purpose of this study Types 10, 20, 22, 25, and 26 have been designated as probable memory questions, the remaining types as probable thought questions. The percent of each group is given in the last two lines of Table VIII. These percents can be considered as only a rough indication of the relative frequency of these two very general divisions.

In her investigation of "the question as a measure of efficiency in instruction," Dr. Stevens¹ attempted to determine the relative number of thought questions and memory questions asked by teachers in a single class period. The percents of memory questions for history, English and science were 83, 55, and 67 respectively. This relative frequency is much larger than indicated in Table VIII. The difference may be due to the fact that in the present investigation only written examination questions were considered, but it is altogether likely that it is indicative of a real change in the type of questions which teachers commonly ask of their students.

Relation of questions to educational objectives. The questions which teachers ask during class periods constitute a concrete expression of the educational objectives which they are day by day setting for their students. The questions of the final examinations should, therefore, be representative of the types of educational objectives set in the different school subjects.² The emphasis upon memory and some of the simpler types suggests a need for a modification in emphasis in most of the school subjects.

Quality of examination questions. Altho the writers have no objective evidence to present in regard to the quality of examination questions, those submitted for this study were in general considered good. Catch questions or those stated so that they would not be understood easily by students were very rare. Many questions were stated so that the grading of the answers was objective and would indicate that their form had been influenced by

¹Stevens, Romiett. "The question as a measure of efficiency in instruction." Teachers College Contributions to Education, No. 48. New York: Teachers College, Columbia University, 1912.

²See page 55 for a further discussion of objectives in school subjects.

the exercises of standardized tests. This was especially true of those examinations having questions of the true-false or recognition type. In the judgment of the writers the criticism that teachers are inclined to ask misleading or catch questions is not a valid one in the case of the examinations studied in this investigation.

CHAPTER VI

THE IMPROVEMENT OF WRITTEN EXAMINATIONS

Altho we now have a number of standardized tests which are superior to written examinations, and we have reason to believe that they will be used even more extensively than at present, there is need to give attention to the improvement of written examinations. It does not appear likely that standardized tests will ever replace written examinations. The latter type of measuring instrument will probably continue to be the most frequently used means of measuring the achievements of school children.

Written examinations may be improved by correcting the faults which have been noted in the preceding chapter. In this chapter we shall consider four important improvements: (1) Reduction of constant errors; (2) Reduction of variable errors; (3) Securing a greater agreement of the content of examinations with recognized educational objectives; (4) Simplification of the administration of written examinations.

There is some overlapping between these improvements. For example, the magnitude of errors in measurement, particularly variable errors of measurement, will be reduced by securing a greater agreement between the content of the examination and recognized educational objectives. The devices for simplifying the administration of examinations also tend to make the results more accurate.

Causes of constant errors in examination grades. The fundamental cause of constant errors in examination grades, i.e., "high grades" or "low grades," is the failure to recognize the distinction between "scores" and "grades." (See page 11.) A pupil's grade tells his standing with reference to a norm, i.e., the passing mark. When no distinction is made between "scores" and "grades" this norm is subjective. Altho the passing mark may be defined numerically as 70 percent or 85 percent it is fixed in the case of a particular examination by the difficulty of the questions and by the severity of the marking of the papers. Pupils will receive "high grades" when the examination is easy or the plan of

marking is generous. They will receive "low grades" when the examination is hard and a severe plan of marking is followed. If the teacher makes no distinction between "scores" and "grades" he sets the norm for a particular examination when he makes out the questions and decides upon the plan of marking. He implicitly expresses the opinion that the pupil whose achievements are barely "passing" will make a grade of 70, or the passing mark adopted. He also implies that the pupil whose achievements are exceptionally high will make a high grade, i.e., a grade of 95 or between 95 and 100. Such expressions are merely subjective.

Since the failure to recognize the distinction between "scores" and "grades" is the cause of constant errors the plan for improvement is obvious. The papers should be marked in terms of "scores." These may be on the scale of 100 but this is not essential. In fact it will probably assist a teacher in keeping the distinction in mind if the scores are not on the scale of 100. After the papers have been marked the "scores" should be translated into "grades" by comparison with a norm in which the subjective elements are reduced to a minimum.

A standard average grade used as a norm. The simplest objective¹ norm is a standard average grade. This may be set arbitrarily but a more rational procedure would be to take the average of the grades given in a school on a particular subject during a period of several years.

The standard average grade defines the grade into which the average score of a typical class should be translated. For example, if the standard average grade is 85 and the average score in a particular class is 57 the grade corresponding to this score would be 85. In case the class is made up of poor students the average grade of the class should be below the standard average grade. If the class is unusually bright their average grade should be higher than the standard average. The translation of the average score into the appropriate corresponding grade furnishes a basis for the translation of the other scores of the group.

The procedure just outlined is necessarily crude. It is partially subjective because the determination of the general status of

¹The adjective "objective" is not intended to indicate perfect objectivity or even as high a degree of objectivity as we have in the case of many standardized tests. As used here it means that the norm is distinctly less subjective than the norm commonly implied in the usual examination.

the class is left to the teacher. However, the teacher may use previous school records or the measures obtained from a standardized test to assist him in arriving at a partially objective estimate of the general status of the class. The use of a standard distribution instead of merely a standard average grade represents a more systematic procedure.

1. Decreasing the magnitude of constant errors by means of a standard distribution of grades. For several years a number of educators have been urging that teachers make the distributions of their grades conform to a standard shape, i.e., that a specified percent of the members of a typical class be given a grade of A, another specified percent a grade of E, and so on for each of the marks adopted by the school.² A number of distributions have been recommended. For a five point system of grades several authors have recommended the following distribution, 7, 24, 38, 24, 7. Other distributions which have been advocated are 7, 18, 50, 18, 7 and 5, 15, 60, 15, 5.

The essential feature of the plan is a specification of the percent of the students of a given group who are to receive each mark rather than the particular form of distribution used. There is much evidence which indicates that the distribution of achievements of an unselected group of students approximates the normal probability curve.³ If we assume that true measures of the achievements of an unselected group of 100 or more are distributed normally this adjustment does not fix the percent who are to receive each grade. The normal probability curve may be divided in many ways, for example, it is possible to divide the curve so that there would be 50 percent of A's, 20 percent of B's, 10 percent of C's, 10 percent of D's, and 10 percent of E's. In such a distribution a grade of A would be given to all students above the average of the class. An appropriate meaning could be stated also for each of the grades. A distribution which is symmetrical has certain advantages and one of those mentioned in the preceding paragraph is to be preferred. The particular standard distribution to be used is a matter of policy which each school should determine. Some argue that different standard distributions be adopted for

²If grades are expressed in percents the corresponding intervals such as 95 to 100, 90 to 94, etc. would be used instead of A, B, C, etc.

³The normal probability curve is bell shaped and is symmetrical with the average or median as a center.

the different years of the high school, some advocate different standard distributions for different school subjects. It should be noted, however, that there are certain advantages in uniformity. It would be desirable for all high schools, particularly those in a given state, to agree upon a common standard distribution and to use this for all subjects. Grades assigned in different schools can have a common meaning only when they conform to the same standard distribution.

The proposal that teachers make the distributions of their grades conform to a standard shape has met with much criticism. As in any controversy there have been extremists on both sides and many of those participating have given evidence that they failed to understand clearly the nature of the proposal of its essential features. Among the advocates of the use of a standard distribution are those who have insisted that the normal probability curve explicitly defines the students who must receive A's, who must receive B's, etc. Cases have been reported of instructors who frankly admitted that a certain student deserved to receive an A but that they had used up all the A's which the distribution allowed, and that, therefore, the student must be satisfied with a B. One hears also of instructors who announce at the beginning of a course that a certain number of the class must fail. It is rumored that in some of these instances the students enrolled have hired certain other students who were indifferent to their scholastic standing to enter the course in order to provide the requisite number of failures. The opponents of the use of a standard distribution have contended that there was no a priori reason why any student should fail and that always the quality of the student's work should determine his scholastic standing. Furthermore, they have pointed out that in any group of students brought together for instructional purposes it is extremely unlikely that the distribution of achievement would approximate at all closely their predetermined standard distribution. The mechanical and unintelligent application of a standard distribution by some instructors has given the opponents of the plan concrete examples of what they imagined to be the normal result of its use.

A standard distribution is merely a device which teachers may use in order to reduce to a minimum the constant errors in their grades, but to be helpful it must be used intelligently. It must be remembered that a standard distribution is a means and

not an end. Whenever common sense indicates that the distribution of grades for a particular class should depart from the normal distribution no instructor should hesitate to award the grades which he believes the students deserve. It is intended that the standard distribution will be closely approximated only for a large unselected group of students. A particular class very frequently is made up of a selected group of students. Furthermore, classes of the usual size, 20 to 35, are so small that frequently there will be significant departures from this standard distribution.

Translating scores into school marks by means of a standard distribution. A standard distribution is useful in translating examination "scores" into "grades." The examination papers should be marked in terms of a score. This score may or may not be on the scale of 100 points. In order to avoid confusion between "scores" and "grades" it is wise to use a scale of points shown so that the maximum score will not be 100. If the papers have been

74	} A	marked in this way the scores may be arranged in columns as indicated in the left hand margin. The first step in translating these scores into grades is to determine whether or not the class is typical. If an experienced teacher has had a class for several weeks he will usually be able to estimate its general status with a fair degree of accuracy. At the beginning of a school year or in the case of an inexperienced teacher some outside information is needed. The previous school record of the students may be studied but in many cases it will be more convenient to administer a general intelligence test. The average mental age and the distribution of the I.Q.'s of the class will be a very reliable index of the composition of the group. If the median I.Q. is distinctly below 100 the teacher may know that he has poor pupil material. If it is much above 100 he knows that the class consists of pupils better than the average. If there is an unusually high number of low I.Q.'s he may expect a relatively high number of low grades.
73		
70		
69		
68	} B	
64		
63		
60		
58	} C	
57		
56		
55		
54		
51		
50	} D	
47		
46		
42		
41		
40	} E	
38		
35		
32		

With the general status of the class in mind the scores may be grouped in conformity with the system of marks used. In the illustration in the left hand margin it has been assumed that the class is approximately typical. The percent of A's and also the percent of failures

are somewhat larger than the percent specified in most standard distributions. If the scores are arranged in the form shown below the general shape of the distribution will be more obvious. However, in the majority of cases it will be sufficient to use the arrangement given in the margin.

			58	
			57	
	47	56	69	
	46	55	68	
38	42	54	64	74
35	41	51	63	73
32	40	50	60	70
E	D	C	B	A

It is seldom that one will have exactly a symmetrical distribution of grades for a class of this size. Some departures from the standard distribution must be expected. In case the class is not typical one should expect marked departure from the standard distribution. For example, the distribution for a given class might be as shown below. In this there are no grades below passing but there are a number of poor students just above the passing mark. Also the percent of A's and B's is unusually large. Such a distribution is not normal but might well represent the distribution of grades for a particular class even when the normal distribution had been adopted as the standard. If the teacher is able to show that the general status of the class justifies such a departure he deserves commendation rather than criticism for his distribution.

			64	
			63	
	42			
	41	54	60	74
	40	51	58	73
	38	50	57	70
	35	47	56	69
	32	46	55	68
E	D	C	B	A

An accumulative distribution used as a check upon constant errors. A standard distribution is also useful as a check upon the grades given by a teacher over a period of several terms. When the grades for the entire period are assembled in such a distribution any general tendency on the part of the teacher to give too high or too low grades will be revealed. Each teacher should keep an

accumulative distribution of the grades in each subject he teaches. For example, a teacher in mathematics should keep an accumulative distribution of the grades given in classes of first-year algebra. When the total number of grades becomes large a comparison of this distribution with the standard distribution will reveal any tendency on the part of the teacher to grade too high or too low in this subject. A teacher should then take steps to correct any marked departures from the practise defined by the standard distribution. In large schools where there are several sections of the same subject it will be helpful to secure a distribution of grades each time they are issued. Any marked departures from the standard distribution will then be called to the attention of the teachers. However, one should avoid giving the impression that there must be uniformity with the standard distribution. Departures from this standard distribution are justified when the group of pupils can be shown to be selected. Thus a departure from the standard distribution is a cause for an investigation on the part of the teachers concerned. If evidence can be produced which justifies the departure no change in the system of grading should be used. On the other hand when investigation reveals no reasons why there should be departures from the standard distribution, the teachers should be urged to modify their system of grading so that a greater uniformity will be secured.

2. Decreasing the variable errors in examination scores.

The reduction of the magnitude of variable errors of measurement in examination scores is to be secured mainly through the adoption of rules which will bring about greater uniformity in preparing and administering examinations. These rules should include specifications in regard to the effect of poor writing, poor spelling, and poor English upon a student's grade, and should be in agreement in regard to giving credit for correct principle and partial credit for exercises partly right or partly completed. The rules may properly include also specifications relating to the number and types of questions to be asked and the form in which they are to be presented to the students. For guidance in marking papers a teacher should write out, at least in abbreviated form, the correct answers to the questions. The accuracy of examination scores will be increased also by making the examinations more uniform with respect to content.⁴

⁴For recommended rules covering these and other points see Chapter VII.

It has been proposed that the use of types of questions which call for answers that may be objectively classified as either "right" or "wrong," would facilitate uniformity in marking the papers. This means of reducing the variable errors of measurement will be considered under the head of "simplifying the administration of written examinations."

3. Securing agreement of the content of examinations with recognized educational objectives. The intrinsic function of an examination is to measure certain achievements. In general the achievements for which we desire to secure measurements are those included in the recognized educational objectives. Hence, the questions should be in agreement with the objectives. Therefore, it is impossible to cover all details in a given subject-matter field. The questions should relate to the most significant facts, principles, etc. of the course. Catch questions and those calling for unimportant details have no place in an examination. For example, an examination in spelling should not include unusual or obsolete words, an examination in history should not call for obscure dates or other trivial facts.

In securing agreement the teacher should make use of such terms of minimum essentials as are available. For example, in spelling a teacher may very properly select the test words from Ayres' list of the one thousand most frequently used words or from some other carefully prepared minimum essential list. In geography a teacher will find the Hahn-Lackey Geography Scale a helpful source of questionings. In other subjects the teacher will not have access to terms of minimum essentials as complete as in these two subjects, but he should become familiar with curriculum studies and other investigations⁵ relating to educational objectives.

⁵The following list is suggestive of studies relating to educational objectives: Yearbooks of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company.

Part I of 14th—reading, writing, spelling, language and grammar, arithmetic, history, literature, geography.

Part I of 16th—reading, writing, spelling, arithmetic, history, physical education.

Part I of 17th—arithmetic, geography, reading, English, civics, history.

Part II of 17th—history, civics, economics, sociology, geography.

Part I of 19th—on new materials of instruction, reading, history, geography, mathematics, nature study, civics.

Part I of 20th—on materials of instruction—all subjects in elementary schools.

Part II of 22nd—the social studies in the elementary and secondary school.

"Arithmetic, course of study for the elementary schools, including the kindergarten and the first six grades." Course of Study Monographs, Elementary Schools, No. 1 Berkeley, California: Public Schools, 1921. 86p. (*Concluded on p. 56.*)

In some subjects there are valuable committee reports which give the consensus of opinion concerning the relative importance of the numerous distributions.

The teacher must assume most of the responsibility for securing the agreement between the content of the examination and educational objectives. In many of the high-school subjects he can obtain little assistance from such sources as just indicated. However, if this purpose is kept in mind and if he is really familiar with the subject which he is teaching, gross inconsistencies with recognized educational objectives will be avoided.

4. Simplifying the administration of written examinations.

The administration of written examinations, particularly the marking of the papers, can be greatly simplified by the use of certain types of exercises. For example, in the true-false type of exercise the pupil merely indicates whether the statement is true or false. Instead of asking the question, "Why did the Puritans come to America in the seventeenth century?" we may ask whether the following statement is true or false. "The Puritans came to America in the seventeenth century seeking wealth." The pupil may give his answer to this exercise by writing a plus sign after the statement if he considers it true and a minus sign if he considers it false. In case the statement is dictated to him he may write after the number of the exercise the word "true" or "false" or the appropriate sign. The answering of such exercises requires very little of the pupil's time and the scoring is exceedingly simple. Questions which can be answered merely by "yes" or "no" also simplify the administration of examinations. Similar results can be secured with recognition exercises such as have been used in a number of standardized silent reading tests. The following is an exercise of this type.

Ayres, L. P. "A measuring scale for ability in spelling." N. Y.: Division of Education, Russell Sage Foundation, 1915. 58p.

Ayres, L. P. "Measuring scale for handwriting." N. Y.: Division of Education, Russell Sage Foundation, 1920. (Folder, chart.)

Bagley, W. C. and Rugg, H. O. "The content of American history as taught in the seventh and eighth grades." University of Illinois Bulletin, Vol. 13, No. 51. Urbana: University of Illinois.

Charters, W. W. Curriculum Construction. N. Y.: Macmillan Co., 1923. 352p.

Charters, W. W. and Miller, Edith. "A course in grammar." University of Missouri Bulletin, Vol. I, Education Series 9. Columbus: University of Missouri, 1915.

Hahn, H. H. Hahn-Lackey Geography Scale. Wayne, Nebraska: H. H. Hahn, State Normal School.

Hahn, H. H. Scale for Measuring Ability of Children in History. Wayne, Nebraska: H. H. Hahn, State Normal School.

Moore, E. C. Minimum Course of Study. N. Y.: Macmillan, 1923. 402p.

“The first president of the United States was: Christopher Columbus, Benjamin Franklin, George Washington, Thomas Jefferson.”

In answering this exercise the pupil is asked to underline or mark in some other way the name required to make a true sentence. Completion exercises in which pupils are asked to supply words which have been omitted furnish still another means of simplification.

Directions for constructing a true-false examination.⁶ 1. In constructing true-false exercises, a list of statements covering in some detail the portion of the subject on which the pupils are to be examined should be prepared. Some of the statements can then easily be changed so that they are false. The untruth of a statement should not be too obvious or it will be worthless for testing. Also statements should be selected which require an acquaintance with the subject in order to determine their truth or falsity.

2. In a true-false examination the number of true statements should approximate the number of false statements, and the arrangement should be such that there is no regular sequence between true statements and false statements.

3. Since the pupil can give his responses very quickly, the examination should consist of not less than fifty statements. A true-false examination of one hundred statements can be given in the time usually devoted to an ordinary examination.

4. The examination should be mimeographed or printed so that each pupil will have a copy. He may give his answers in the margins of the sheets, or, if it is desired to use the same set of papers with another group of pupils, he may be given a sheet of paper on which there are numbered blanks. The pupils will then be asked to record in the blanks their answers to the corresponding exercises. A less desirable plan, which may be followed when it is not possible to secure mimeographed copies of the examination, is to read the statements to the pupils and have them record their answers in numbered blanks. The disadvantage of this plan is that the pupils do not have a satisfactory opportunity to study the statements. Also the class may give some indication of the answer if a statement appeals to them as being ridiculous.

5. The pupils should be given specific directions in regard to answering exercises about which they are uncertain. One writer

⁶For an example of a true-false examination, see Appendix p. 69.

has suggested that the pupils be instructed to guess concerning the truth or falsity of the statement. Another writer who has used this type of examination instructed the pupils as follows: "First, go through the list quickly and mark all that you know for certain, then go back and study out the harder ones. Do not guess; the chances are against you on guessing. Don't endanger your score by gambling on those questions about which you know nothing." This second procedure is probably the better.

The scoring of a true-false examination. Since only two responses are possible, it is obvious that a pupil may give a correct response as the result of chance. In order to take this possibility into account, a pupil's score on an examination of this type is the number of exercises answered correctly minus the number answered incorrectly. Exercises not attempted are not counted.

Directions for constructing a recognition examination.⁷ In constructing this type of examination none of the proposed answers should be too obviously incorrect. An exercise can yield an indication of a pupil's achievement only when he is forced to use judgment in determining which of the proposed answers is suitable. For example, the illustrative exercise given would be practically worthless for testing purposes if all the names, except that of George Washington, were of persons living today or of persons having no connection with our national life. In applying this type of exercise to the field of arithmetic the proposed answers should include erroneous answers which pupils are inclined to give: if the exercise called for the quotient of two fractions, one of the proposed answers should be the product of the fractions and another their sum, and perhaps another should be the fraction obtained by taking the sum of the numerators as a new numerator and the sum of the denominators for a new denominator. When the correct answer is included in a group of such answers as these, the pupil who does not know how to find the quotient of such fractions will be unable to determine the correct answer except as a matter of chance. On the other hand, if all of the answers except the correct one were integers or were so large that they were obviously incorrect, a bright pupil who knew nothing about division of fractions would be able to select the correct answer. The correct answer should not always be found in the same position; sometimes it should be

⁷For an example of a recognition examination see Appendix p. 75.

first, sometimes last, and sometimes in an intermediate position. As in the case of the true-false examination, a recognition examination should consist of a large number of exercises.

Examinations of this type should be mimeographed or printed and each pupil should have a copy. Definite instructions concerning methods of work should be given. It is probably best to instruct the pupil to work through the test rapidly, answering those exercises about which he is certain. He should then go back over the list and try the more difficult ones. Not fewer than four proposed answers should be included in each statement and the pupils may be instructed to guess if they do not know, since the chance of success by guessing is slight. The pupil's score on an examination of this type may be taken as the number of exercises done correctly.

A somewhat unusual but interesting type of recognition exercise is that described as a "matching contest." In this a pupil is given two lists of statements, the first numbered 1, 2, 3, 4, 5, etc., the second marked A, B, C, D, E, etc. In the second list, there is a statement which corresponds in meaning to a statement in the first list and the pupil is to pair these statements, marking by the number of the first list the letter of the corresponding statement of the second. For example, in the exercises given below: by the date marked (5), 1898, we place the letter B to indicate the event for which that date is significant. It is difficult to construct such examinations so that they will require reasoning on the part of the student. Their most important use is in the elementary school for rapid drill in certain phases of some subjects, such as definitions in geography and grammar, etc. The following exercises, selected from the Spokane United States History Test, illustrate the use of such an examination in linking a certain date or person with the corresponding event.

1. 1846	A. Lincoln's Emancipation Proclamation
2. 1865	B. Spanish-American War
3. 1863	C. Beginning of World War
4. 1917	D. Declaration of Independence
5. 1898	E. United States entered World War
6. 1789	F. Election of Washington as President
7. 1792	G. War with Mexico began
8. 1776	H. Invention of the cotton gin
9. 1861	I. Lee's surrender at Appomattox
10. 1914	J. Beginning of Civil War

1. Foch	A. Destroyed Spanish fleet in Manila Bay
2. Lincoln	B. Invented the telephone
3. Fulton	C. Leading Confederate General
4. Dewey	D. Wrote the Declaration of Independence
5. Pershing	E. Invented the steamboat
6. Bell	F. Commanded allied armies in the World War
7. Edison	G. Was President during the Civil War
8. Jefferson	H. Commanded American Forces in the World War
9. Lee	I. Was Revolutionary patriot, author, and inventor
10. Franklin	J. America's most famous inventor

Directions for constructing completion exercises.⁸ A completion exercise should be constructed so that no suggestion will be given of the correct words to be written in the blanks. Furthermore, the facts to be supplied should be important. The best plan is to prepare a list of important statements and principles covering the portion of the subject over which the pupils are to be examined and then from these statements to strike out a certain significant word or phrase. In every case, if it is possible, the words omitted should be such that only one answer will be correct. Since little writing is required of the pupils they may be asked to fill in as many as one hundred blanks.

The scoring of completion exercises is not as highly objective as in the two types mentioned above. Pupils will tend to write a variety of words in the blanks. Different words may have almost the same meaning, and frequently the scorer will be compelled to determine whether the meaning of some word is sufficiently near that of the correct answer to justify giving the pupil credit for having answered the exercises correctly. However, by a careful selection of statements and of the omitted words, this subjectivity may be greatly minimized. For example, in the sentence, "The first Continental Congress was held in . . .," only one possible word can be correct. In using completion exercises it is necessary to provide each pupil with a mimeographed or printed copy of the examination. The pupil's score is the number of blanks filled in correctly.

Advantages of the "new examination." Examinations consisting of exercises of the types described above have certain obvious advantages. There will be a large saving of time for both teacher and pupil. The pupil is called upon to do little or no

⁸For an example of a completion examination see Appendix p. 73.

writing in giving his answers and therefore is able to respond to a large number of exercises. The teacher in scoring will have little or no occasion to use judgment as he will need only to note the brief responses given by the pupils. Thus the labor of scoring will be greatly reduced and, more important, the scoring will be much more highly objective than that in the marking of examination papers of the usual type. The saving of time in the giving and scoring of the "new examination" will more than offset any additional time that may be expended in its construction. Another advantage is that the new examination can be made more comprehensive. Examinations as a rule consist of ten questions. Some are limited to a smaller number. Consequently the scope of examinations of the traditional type is necessarily narrow. "New examinations" of the true-false type should consist of not less than fifty exercises and may have as many as one hundred. Other types of the "new examination" should be of a corresponding length. Hence a "new examination" will usually be more comprehensive than a traditional examination.

Limitations of the "new examination." There are certain limitations of the new examination which should be noted along with its advantages. It can not be used in mathematics except to a limited extent. It can not be used at all in English Composition. In other subjects there are many phases of achievement which are not measured directly by examinations made up of exercises of the types described above. Hence, altho the "new examination" is more comprehensive with reference to information, and does measure certain types of achievements, it is likely that pupils would miss much valuable experience and training if they were not at times asked to "compare," "explain," "discuss," "define," or "tell why." They should also be asked to summarize material presented on a topic or to apply certain principles. The following questions taken from Hahn's Scale for Measuring Ability of Children in History appear to require mental processes distinctly different from those called for by the "new examination."

1. "State points of similarity between the position of the United States in 1812 and their position in 1912."
2. "Arrange the following events in order of cause and effect: Force Bill, Carpet Baggers, Fifteenth Amendment, Negro Rule in Some of the Southern States, Ku Klux Klan."
3. "Name the presidents of the United States since 1892."

An intelligent attitude toward the "new examination." The simple administration of the new examination and other attractive features should not blind one to the limitations just mentioned. As indicated in Chapter II written examinations do more than merely secure measures of achievement. If they consist of the right kind of exercises they afford significant educational opportunities. The educational opportunities of the "new examination" are necessarily restricted, and it would be unfortunate if it entirely replaced examinations of the traditional type. The new examination, however, has a place. It may be used occasionally in most school subjects. It is useful when a teacher wishes to test the acquaintance of a class with a wide range of facts. It has little diagnostic value and examinations of the traditional type should be used when information is desired concerning the weaknesses of different members of a class. For this reason the "new examination" is more appropriate for use at the end of a term than for tests during the term which have as their purpose both measurement and diagnosis.

CHAPTER VII

RULES FOR THE PREPARATION AND ADMINISTRATION OF WRITTEN EXAMINATIONS.

Below, a group of suggested rules governing the preparation and administration of written examinations are given. These represent the opinion of the writers which is based upon a careful study of the problems involved, as well as upon several years of experience in the measurement of school achievement.

1. Final examinations should be required. In school subjects such as shop work, in which the performances secured from pupils are highly objective, the waiving of this requirement may be justified. When final examinations are given no student should be excused from them because of high daily grades, deportment or attendance. (See p. 16)

2. The content of final examinations should agree as closely as possible with recognized educational objectives. In fields where minimum essentials have been determined they should be used as a basis in formulating questions. (See p. 55)

3. The questions should be definite and stated so that all pupils will interpret them alike. Questions relating to items of minor importance should occupy a minor place in examinations. Questions relating to points which have not received attention in the course should be omitted.¹ (See also rule 11.)

4. When the necessary equipment is available the questions should be mimeographed or typewritten so that each student will have a copy on his desk. In case they are written on the board the teacher should make certain that all pupils are able to read them correctly. It is well in either case to read the questions aloud to the class.

5. The examination should be sufficiently difficult so that few pupils will make perfect scores. (This rule should not apply

¹Frequently in an examination the difficulty of the question is due to the lack of emphasis placed upon it throughout the term because it deals with a relatively unimportant topic. Other topics, difficult in themselves, but emphasized because of their importance, furnish the easier questions.

when the purpose of the examination is to determine which students have attained a given standing which includes perfection of performance.) (See p. 12)

6. Usually the examination should be long enough so that every member of the class will be kept busy for the entire period. It is better to make the examination too long than to have it too short. In this way it becomes possible also to take into account the student's rate of work in determining his grade. Appropriate adjustments can be made in interpreting "scores" into school "marks." (See p. 22 and 13)

7. In questions asking for a discussion or explanation indicate the completeness of the discussion or the degree of elaborateness expected in the answer. (See p. 24)

8. Time may be economized for both students and teacher by using some form of the "new examination." This type of measuring instrument, however, possesses certain limitations which should be kept in mind. The exclusive use of it would be unwise. (See p. 61)

9. Unless the students have a definite understanding of the methods of work which are to be followed, the teacher should give them explicit directions concerning such matters as the order in which the questions are to be answered, the desired arrangement of the work, and any other items in which there is an opportunity for pupils to adopt different procedures. (In the case of most standardized educational tests, the directions to students are very detailed and explicit.) (See p. 24)

10. Approximately ninety minutes should be allowed for a final examination in most high-school subjects. The time which teachers should devote to the preparation of the questions will depend upon their experience and upon their practise during the semester. It is recommended that a teacher make a record throughout the term of questions which in his judgment are suitable for a final examination. From two to three hours should be sufficient for marking a set of twenty-five examination papers. If a teacher finds that a longer time is required, he should endeavor to modify his procedure so that this work can be done more quickly. (See p. 19)

11. Altho any weighting of questions by a teacher will be subjective, it is probably desirable to weight the questions, particularly in cases of extreme differences in value. The weighting,

however, should be upon the basis of social importance rather than upon mere difficulty. (See p. 22)

2. It is advisable for the teacher to write out at least in an abbreviated form the answers to examination questions before he begins marking the papers. In mathematics and other subjects where a definite answer is required and only one can be accepted as correct, the need for this rule is not as great as in such subjects as geography, history, literature and certain phases of science. However, a list of correct answers will usually mean a saving of time. (See p. 23)

13. Except in courses in English a pupil's grade should not be intentionally lowered for errors in spelling or for poor handwriting. As a rule the grade should not be lowered because of poor English, unless the quality of the English is evidence of unsatisfactory reasoning. Rules covering these points, as well as others concerning which teachers might differ, should be formulated by the principal in conference with his teachers or at least a committee of them. The rules thus formulated should be carefully followed by all of the teachers. (See p. 20)

14. In marking the papers more accurate results will in general be secured if the answers to one question are marked on all the papers before those for another question are taken up. (See p. 23) When it is desired to mark all of the questions on one paper before taking up another, the "sorting method" should be used. According to this procedure the papers as they are read are sorted into piles, the best ones being placed in the first pile, the next best in second pile, etc. Five distributions will, in most cases, prove sufficient. After all the papers have been distributed they should be reread, one pile at a time, and compared with each other. If these papers do not possess approximately the same value, changes in the sorting may be made. Grades may then be assigned to the papers in the different piles.

15. The distinction between "scores" and "grades" should be kept in mind. (See p. 11) The papers should be marked first in terms of scores. In doing this an appropriate number of points should be determined for each question. It is not necessary that the total of these points be 100. (See p. 52)

16. The point scores assigned to the examination papers should be translated into school marks. In doing this the use of a standard distribution will be found helpful, and will operate also to decrease the magnitude of the constant errors. (See p. 50)

APPENDIX

(Questionnaire Sent to High-School Teachers)

QUESTIONNAIRE RELATING TO THE USE OF WRITTEN EXAMINATIONS IN HIGH SCHOOLS

Teacher.....High School.....City.....

The major subject which I am teaching is.....

The following questions are to be answered only with reference to the major subject you are teaching.

1. Approximately how much time do you use in preparing questions for a final examination which the students are allowed a total of 90 minutes to answer?.....Min.

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester.)

2. In preparing a set of examination questions do you usually attempt to arrange these questions in order of ascending difficulty?..... Yes No*

3. Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answers, arrangement of work, etc.) Yes No

4. Which of the following methods do you use in presenting questions to the student?.....

(a) Writing the questions on the board..... Yes No

(b) Furnishing the pupils with a mimeographed, carbon or printed copy of the questions..... Yes No

(c) Dictating the questions to the pupils..... Yes No

5. Is it your custom to make the examination long enough so that practically none of the students will answer all of the questions in the time allowed?..... Yes No

*Underline the answer (Yes-No) which you desire to make.

6. Is it your custom to make the examination short enough so that practically all of the students will answer all of the questions in the time allowed?..... Yes No
7. If you have given an affirmative answer to Question 6, do you note the time each student spends in writing his answers?..... Yes No
8. What proportion of the final mark for the semester is based upon the final written examination grade?.....%
9. In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?..... Yes No
10. Do you usually grade all the answers on one paper before taking up those of another paper?..... Yes No
11. Do you usually grade the answers to one question on all of the papers before taking up the answers to a second question?..... Yes No
12. Instead of marking the answers to each question separately do you attempt to estimate the value of the paper as a whole?..... Yes No
13. Before starting to grade a set of examination papers do you write out the answers which you consider correct?..... Yes No
14. When you consider the questions of an examination to be unequal in difficulty is it your practise to give more credit for a correct answer to a difficult question than for a correct answer to an easy question? Yes No
15. Approximately how much time do you use in marking the papers of a final examination which the students are allowed a total of 90 minutes for answering? Estimate as accurately as possible. Base this answer on a class of 25 students.....Min.
16. In marking examination papers do you intentionally lower a student's mark in the case of
- (a) poor writing..... Yes No
- (b) poor spelling..... Yes No
- (c) poor English..... Yes No
17. In the case of questions which are essentially mathematical in character do you give credit to the student for using the correct principle even though the final answer be wrong?..... Yes No

NOTE: It is desired that only teachers of mathematics, physics, and chemistry answer No. 17.

NOTE TO TEACHER: When you have answered the above questions please return this questionnaire to your principal.

(Questionnaire Sent to High-School Principals)

Principal.....High.....City.....

1. Do you require your teachers to give final examinations at the end of each semester?..... Yes No*

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester)

If a *negative* answer is given to the first question, no answer is expected for the remaining questions.

2. a. Is it the practise in your school to exempt certain students from final examinations?..... Yes No

b. If so, what are the conditions (requirements) upon which you base exemption?

1. Deportment..... Yes No

2. Scholarship..... Yes No

Other requirements.

3. How many minutes do you allow for final written examinations?....Min.

4. Because final examinations have proven unreliable some educators urge that students be given more than one comprehensive examination in each subject, and that these examinations be given on different days. Do you require more than one such final written examination in each subject?..... Yes No

5. a. In marking examination papers is it the practise in your school for the teachers to subtract from a pupil's grade for

1. poor writing..... Yes No

2. poor spelling..... Yes No

3. poor English..... Yes No

b. Are your teachers accustomed to giving more credit for correct answers to difficult questions than for correct answers to easy questions?..... Yes No

c. Is it the practise of your teachers when computing a semester mark to add or deduct credit in proportion to the time used by a pupil in answering the final written examination questions? Yes No

6. a. Have you advised your teachers as to what proportion of the final mark for the semester should be based on the final written examination?..... Yes No

b. Have you made a definite requirement in this respect?..... Yes No

c. If so, what proportion of the final mark for the semester do you require to be based upon the final written examination mark?..%

7. Additional information pertaining to this topic is called for in a second questionnaire which is to be filled out by high school teachers. Would you be willing to distribute this questionnaire among your teachers? If so, will you kindly indicate the number of teachers in your high school?.....

*Underline the answer (Yes-No) which you desire to make.

EXAMPLES OF "NEW EXAMINATIONS"

(The following "new examinations" are given for purposes of illustration. They may include several exercises which will prove unsuitable when given to pupils.)

TRUE-FALSE EXAMINATION IN PHYSIOLOGY

Prepared by

Bureau of Educational Research

University of Illinois

Name.....Boy or Girl.....

Age last birthday.....Next birthday will be.....19.....

Grade.....Date.....City.....State.....

School.....Teacher.....

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (-) in the column to the right.

EXAMPLES

Read the statement below very carefully.

1. **Fats will form a lasting mixture with water.**

This is not a true statement so you will place a minus (-) sign in the column. Now read the second sentence.

2. **The layer of fat just beneath the skin is more than one-tenth of an inch thick.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the third sentence.

3. **The union of oxygen with any substance produces heat.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the fourth sentence.

4. **Nitrogen constitutes only one-fifth of the volume of the air.**

This is a false statement so you will mark a minus (-) sign in the column.....

Answers to be
written here

PHYSIOLOGY

Answers to be
written here

1. The kidneys vary 12 inches to 16 inches in length.
2. The external poisoning of the skin by poison ivy or sumac never results seriously.
3. A person having a good mind must necessarily have a large brain.
4. Color blindness is more prominent in men than women.
5. Plenty of fluids should be drunk at the time of eating solid food.
6. Bones are composed of animal and mineral matter.
7. The nails are hardened outer skin or epidermis.
8. The use of alcohol increases the tendency to commit crime.
9. A full grown person contains about six quarts of blood.
10. The brain is almost perfectly spherical in shape.
11. The kidneys are almost perfectly round.
12. All animals are made up of cells.
13. Substances, like glass, which permit rays of light to pass through them readily are said to be opaque.
14. The sense organs of smell are located in the lining of the cavity of the nose.
15. The skin is composed of two layers of tissue.
16. The end organs for taste occur in the mucous membrane of the tongue.
17. To extinguish the burning clothing of a person, it is necessary to wrap him in something to exclude the air.
18. All of the interior of the spinal cord is filled with gray matter containing nerve cells.
19. The great difference in the complexion of persons is due largely to the pigment lying in the epidermis.
20. Cancer is caused by germs growing in the tissue.

21. Diphtheria can be controlled by the use of Diphtheria antitoxin.
22. The brain is separated into two parts or hemispheres by a great longitudinal fissure.
23. When oxygen is separated from other substances the process is called oxidation.
24. Infectious diseases are due to changed methods of work and growth on the part of cells in certain regions of the body.
25. The use of alcoholic beverages builds up the body and makes the muscles stronger.
26. The great majority of grown people have been infected with tuberculosis germs.
27. The sense organs are the terminations of the sensory nerves serving to carry impressions to the spinal cord or brain.
28. Farsightedness is often caused by a blow on the eye.
29. An antiseptic is a substance which merely restrains the germs from growing.
30. The brain is in communication with the rest of the body by means of nerves.
31. The cerebrum is the path of communication between the nerves supplying the arms, trunk, legs, and brain.
32. The chief function of muscles is to hold up the body.
33. All milk contains bacteria.
34. The alcohol used in drinks is produced by the growth of yeast in a liquid containing sugar.
35. Our blood contains white corpuscles which destroy disease germs.
36. More people die daily from diphtheria than from tuberculosis.
37. The use of tobacco increases the strength of the muscles.
38. The use of tobacco makes the nerve cells function more keenly.
39. The chewing of dry bread aids the digestion as much as the use of gum.

40. Air is composed chiefly of two gases, oxygen and nitrogen.
41. The first step in treating a person who has been poisoned is to give an emetic.
42. Light is produced by waves of a substance called ether.
43. Non-infectious diseases are caused by small plants or animals called parasites feeding upon the human body.
44. Alcoholic beverages have great value in curing disease.
45. A drink of alcoholic beverage in the winter time causes a man's body to become warm.
46. Each portion of the brain has its own definite work to perform.
47. Fainting is caused by an over-sufficient supply of blood being sent to the brain.
48. The spinal cord may act independently of the brain and produce many of the muscular movements necessary in routine work.
49. The germs of typhoid fever usually gain access to the body by being breathed in with air.
50. Narcotics are substances which cause any organs of the body to act more vigorously than is their custom.

Answers to be
written here

Directions to teachers: After the four examples have been studied by the pupils, read the following directions to them: "On the next page you will find a number of statements similar to the ones you have just read. You are to place a plus sign or a minus sign in the column to the right of each statement just as has been done on the first page. Mark all of the statements that you are sure you can answer correctly. If you find a statement that you are not sure you can answer correctly, study it carefully and then mark the answer you think will be correct. If you find a statement you know nothing about, make no attempt to mark it, as guessing counts heavily against you. You will have 25 minutes for the test. I shall expect you to stop promptly and turn your folders face down on the desk when I tell you to do so. Ready-Go."

In computing the score of each pupil on a test subtract the total number of wrong answers from the total number of right answers. Such scores are called "point-scores." In interpreting them it is advisable to form a distribution which will show how many pupils received each score. From the distribution it is possible to work out a basis for translating the point scores into the usual kind of school marks.

NOTE: The "Directions to teachers" given above would not appear on the usual printed examinations. They are placed here for the convenience of teachers.

COMPLETION EXAMINATION IN AMERICAN GOVERNMENT

Prepared by
Bureau of Educational Research
University of Illinois

Name.....Boy or Girl.....
Age last birthday.....Next birthday will be.....19.....
Grade.....Date.....City.....State.....
School.....Teacher.....

Below you will find a number of statements. In each statement one or more important words have been omitted. Each blank in the sentence shows where a word has been left out. Read each statement carefully, then write in the blank the word which completes the meaning of the statement. You will be allowed 15 minutes for the test.

1. The primary purpose for which government exists is the.....of our lives and property.
2. Citizenship may be acquired by.....in this country or by a process offor natives of other lands.
3. Our national government derives its authority from the.....of the United States through our national.....
4. The legislative power granted to the national government is vested in a Congress ofhouses, the smaller of which is called the.....and the larger the.....
5. The execution of the laws made by.....is intrusted to the.....of the United States.
6. All judges connected with the national courts are appointed for life with the consent of.....
7. Most of the candidates for office which are filled by popular vote are nominated directly in.....
8. The Fifteenth Amendment of the United States Constitution prevents the states from denying to citizens the right to vote on account of.....,or previous condition of.....

9. Practically all of our law-making bodies are made up of.....chosen for short terms from.....into which the states, counties and cities are divided.
10. The first permanent English settlements in America were made in what is now the state of.....
11. In a county, the records of the county board and other official papers are preserved by the county.....
12. All cities are public corporations created under.....municipal laws.
13. Every incorporated city obtains from the.....government a.....under which it may elect its officials and conduct its business.
14. Civil service employees may be removed from service only for.....
15. The power of impeaching a state officer is given to the.....
16. The.....is by far the most prominent and powerful executive official in the state. Very.....state officers are appointed by him or are responsible to him.
17. All important officials connected with the executive or judicial service of the United States may be removed by.....through the lower house of Congress and by.....in the senate.
18. Far more property is destroyed by.....than by all other agencies.
19. There is no task of state and local government which outranks in importance that of providing an.....education at public expense.
20. All rivers and canals within a single state are controlled by the.....in which they are located.
21. Most of the revenue for state and local governments is secured by a.....on.....
22. A state.....is the fundamental law which the people of the state have arranged for their government and protection.
23. A state constitution can be changed by means of an.....
24. The three-fifths compromise provided that five.....should be counted as equal to three.....when reckoning the.....for either direct taxation or representation.

RECOGNITION EXAMINATION IN ALGEBRA

Prepared by
Bureau of Educational Research
University of Illinois

Name.....Boy or Girl.....

Age last birthday.....Next birthday will be.....19.....

Grade.....Date.....City.....State.....

School.....Teacher.....

Below you will find a number of statements. In each statement a word or number has been omitted. At the close of the statement several words or numbers have been given. One of these is the correct answer. Select the word or number which you think is correct and draw a line under it. Most, if not all, of the examples can be solved by mental calculation. If any figuring is necessary, work on the margin of the page. You will be allowed 17 minutes for the test.

1. Numbers that are represented by letters are called.....numbers.
substituted—literal
2. When two or more letters are multiplied together each is called a.....
of the product. factor—coefficient
3. If a man rides a certain distance in 10 hours, in h hours he rides.....
 $10h$; $\frac{h}{10}$; $\frac{10}{h}$
4. The statement $2x + 5 = 29$ is called an..... identity—equation.
5. If 16 is subtracted from three times a certain number the result is 110. The number
is..... $36\frac{2}{3}$; $31\frac{1}{3}$; 42
6. A number which is a factor of two or more numbers is called a.....factor.
common—equal
7. If there are two equal factors of a number, either is called the.....of the
number. square root—common factor
8. To multiply algebraic fractions take the.....of the numerators for a
new numerator and the product of the denominators for a new denominator.
sum—product

9. A fraction whose numerator or denominator (or both) contains fractions is called afraction. multiple—complex
10. A.....is a statement of a fact which is to be proved. theorem—axiom
11. The name given the $+$ sign is..... negative—positive
12. To find the sum of two numbers whose signs are opposite, take their.....
regarding each as positive, and prefix the sign of the larger number to the answer.
sum—difference—product
13. Whenever a number occurs without a sign, the.....sign is to be under-
stood. \times ; $+$; $-$
14. The number denoting the power of a term is called the.....
prefix—exponent
15. If $a = 2$, $b = -3$, and $c = -5$ then $\frac{2abc}{a} = \dots\dots\dots -6; -2; 30$
16. In adding like terms add the coefficients for the new coefficient and.....
it by the common factor. multiply—divide
17. An expression which contains more than one term is called a.....
monomial—polynomial
18. If the length of a rectangle $= 4$ feet more than twice the width, the perimeter $=$
56 feet. The length $= \dots\dots\dots$ feet. $8-12-16$
19. Any term may be transposed from one side of an equation to the other, provided
itsis changed. sign—value
20. Any equation which contains no higher power of the unknown letter than the first
is called a.....equation. radical—simple
21. The exponent of the product of two powers of the same number is equal to the
.....of the exponents of the factors. product—sum
22. To raise the product of two numbers to any power, raise the numbers separately
to that power and take their..... product—sum
23. The square of any two numbers is equal to the square of the first number.....
twice the product of the two plus the square of the second number. plus—minus
24. A 20 foot ladder rests against a building, the bottom of the ladder being 12 feet
from the cellar wall. The top is.....feet from the ground. $8-16$
25. In division, the sign of the quotient is.....whenever the dividend and
divisor have like signs. $-; +$

26. In finding the quotient of two powers of the same number the exponent of the quotient is equal to the exponent of the dividend.....by that of the divisor.
increased—diminished
27. $(3x^2 - 2x - 1) \div (x - 1) = \dots\dots\dots 3x + 1; 3x - 1$
28. A factor which has no factor except itself and unity is called a.....factor.
prime—multiple
29. The product of all the common prime factors of two or more numbers or expressions is called their.....common factor. highest—lowest
30. If one number is exactly divisible by another, the first is called a.....of the second. divisor—multiple
31. In algebraic fractions the dividend is called the.....
denominator—numerator

THE
LIBRARY OF THE
MUSEUM OF
ART AND HISTORY
OF THE CITY OF
NEW YORK

UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

Vol. XXI

NOVEMBER 26, 1923

No. 13

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN No. 17

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE PRESENT STATUS OF WRITTEN EXAMINATIONS AND SUGGESTIONS FOR THEIR IMPROVEMENT

By

WALTER S. MONROE
Director, Bureau of Educational Research

Assisted by

LLOYD B. SOUDERS
Formerly Assistant in Bureau of Educational Research



PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

The Bureau of Educational Research was established by act of the Board of Trustees June 1, 1918. It is the purpose of the Bureau to conduct original investigations in the field of education, to summarize and bring to the attention of school people the results of research elsewhere, and to be of service to the schools of the state in other ways.

The results of original investigations carried on by the Bureau of Educational Research are published in the form of bulletins. A complete list of these publications is given on the back cover of this bulletin. At the present time five or six original investigations are reported each year. The accounts of research conducted elsewhere and other communications to the school men of the state are published in the form of educational research circulars. From ten to fifteen of these are issued each year.

The Bureau is a department of the College of Education. Its immediate direction is vested in a Director, who is also an instructor in the College of Education. Under his supervision research is carried on by other members of the Bureau staff and also by graduates who are working on theses. From this point of view the Bureau of Educational Research is a research laboratory for the College of Education.

BUREAU OF EDUCATIONAL RESEARCH
College of Education
University of Illinois, Urbana

THE UNIVERSITY OF ILLINOIS

THE STATE UNIVERSITY

URBANA

DAVID KINLEY, Ph.D., LL.D., *President*

The University Includes the Following Departments

THE GRADUATE SCHOOL

THE COLLEGE OF LIBERAL ARTS AND SCIENCES (Ancient and Modern Languages and Literatures; History, Economics, Political Science, Sociology, Philosophy, Psychology, Education; Mathematics; Astronomy; Geology; Physics; Chemistry; Botany, Bacteriology, Zoology, Entomology; Physiology; Art and Design; Home Economics)

THE COLLEGE OF COMMERCE AND BUSINESS ADMINISTRATION (General Business, Banking, Insurance, Accountancy, Railway Administration, Railway Transportation, Industrial Administration, Foreign Commerce; Courses for Commercial Teachers and Commercial and Civic Secretaries; Commerce and Law)

THE COLLEGE OF ENGINEERING (Architecture; Architectural, Ceramic, Civil, Electrical, Gas, General, Mechanical, Mining, Municipal and Sanitary, Railway Engineering, and Engineering Physics)

THE COLLEGE OF AGRICULTURE (Agronomy; Animal Husbandry; Dairy Husbandry; Farm Mechanics, Farm Organization and Management; Horticulture, Landscape Gardening, and Floriculture; Agricultural Extension; Home Economics)

THE COLLEGE OF LAW (Three-year and four-year curriculums based on two years of college work)

THE COLLEGE OF EDUCATION (General Education; Athletic Coaching; Agricultural Education; Home Economics Education; Industrial Education; Music Education; University High School; Bureau of Educational Research)

THE CURRICULUM IN JOURNALISM

THE CURRICULUMS IN CHEMISTRY AND CHEMICAL ENGINEERING

THE SCHOOL OF RAILWAY ENGINEERING AND ADMINISTRATION

THE SCHOOL OF MUSIC (four-year curriculum)

THE LIBRARY SCHOOL (two-year curriculum for college graduates)

THE COLLEGE OF MEDICINE (in Chicago)

THE COLLEGE OF DENTISTRY (in Chicago)

THE SCHOOL OF PHARMACY (in Chicago; Ph.G. and Ph.C. curriculums)

THE SUMMER SESSION (eight weeks)

EXPERIMENT STATIONS AND SCIENTIFIC BUREAUS: U. S. Agricultural Experiment Station; Engineering Experiment Station; State Natural History Survey; Biological Experiment Station on Illinois River; State Water Survey; State Geological Survey; U. S. Bureau of Mines Experiment Station.

The library collections contain July 1, 1923, 556,105 volumes and 127,941 pamphlets. For catalogs and information address

THE REGISTRAR

Urbana, Illinois

BULLETINS OF THE BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION, UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS

	Price
No. 1. Buckingham, B. R. Bureau of Educational Research, Announcement, 1918-19.....	.15
No. 2. First Annual Report.....	.25
No. 3. Bamesberger, Velda C. Standard Requirements for Memorizing Literary Material.....	.50
No. 4. Holley, Charles E. Mental Tests for School Use. (Out of print).....	.50
No. 5. Monroe, Walter S. Report of Division of Educational Tests for 1919-20.....	.25
No. 6. Monroe, Walter S. The Illinois Examination.....	.50
No. 7. Monroe, Walter S. Types of Learning Required of Pupils in the Seventh and Eighth Grades and in the High School.....	.15
No. 8. Monroe, Walter S. A Critical Study of Certain Silent Reading Tests..	.50
No. 9. Monroe, Walter S. Written Examinations and Their Improvement....	.50
No. 10. Bureau of Educational Research. Relation of Size of Class to School Efficiency.....	.50
No. 11. Monroe, Walter S. Relation of Sectioning a Class to the Effectiveness of Instruction.....	.15
No. 12. Odell, Charles W. The Use of Intelligence Tests as a Basis of School Organization and Instruction.....	.50
No. 13. Monroe, Walter S., and Foster, I. O. The Status of the Social Sciences in the High Schools of the North Central Association.....	.50
No. 14. Monroe, Walter S., and Carter, Ralph E. The Use of Different Types of Thought Questions in Secondary Schools and Their Relative Difficulty for Students.....	.30
No. 15. Monroe, Walter S. The Constant and Variable Errors of Educational Measurements.....	.25
No. 16. Odell, Charles W. An Annotated Bibliography Dealing With the Classification and Instruction of Pupils to Provide for Individual Differences.....	.50
No. 17. Monroe, Walter S., and Souders, Lloyd B. Present Status of Written Examinations and Suggestions for Their Improvement.....	.50

S
,
H

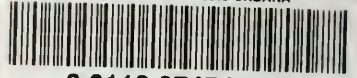








UNIVERSITY OF ILLINOIS-URBANA



3 0112 074742435